DOCUMENT RESUME

ED 212 091 EA 014 383

AUTHOR	Johnstone, James N.
TITLE	Indicators of Education Systems.
INSTITUTION	United Nations Educational, Scientific, and Cultural Organization, Paris (France). International Inst. for Educational Planning.
REPORT NO	ISBN-0-850-38-447-8
PUB DATE	81
NOTE	332p.
AVAILABLE FROM	Publications Officer, International Institute for Educational Planning, 7-9 rue Eugene Delacroix, 75016 Paris, France (Order No. F.73; 100.00 francs, prepaid).
EDRS PRICE	MF01 Plus Postage. PC Not Available from EDRS.
DESCRIPTORS	Comparative Analysis; Diagrams; *Educational Development; Educational Planning; *Educational Quality; Educational Research; Elementary Secondary Education; Input Output; Outcomes of Education; Postsecondary Education; Reliability; *Statistical Data; Tables (Data); Validity
IDENTIFIERS	*Educational Indicators

ABSTRACT

In 11 chapters with 54 data tables and 29 diagrams, the author discusses the theory, problems, and uses of educational indicators. He defines educational indicators as characteristics that point, with more or less exactness, to general or overall aspects of an educational system. Chapter 1 examines the concept of an educational indicator and the considerations to keep in mind when constructing indicators. Chapter 2 presents statistical issues involved with indicators, including their reliability and validity. The next two chapters look at both the theoretical and empirical aspects of developing indicators, while chapters 5 and 6 suggest national and regional indicators of educational inputs, outputs, and processers. The presentation and use of indicators for educational planning and research and for the classification of educational systems are covered in chapters 7-10. In the final chapter the author summarizes the implications of indicator development and use. Three appendices review methodological and statistical aspects of educational indicators. (RW)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED212091

International Institute for Educational Planning

Indicators of Education Systems

James N Johnstone

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☐ This document has been reproduced as received from the person or organization originating it.
☒ Minor changes have been made to improve reproduction quality.

- Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL IN MICROFICHE ONLY HAS BEEN GRANTED BY

J. Hall

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

EA 014 383

Kogan Page, London /UNESCO, Paris

Kepada istri saya Marea yang tercinta

First published 1981 by the
United Nations Educational, Scientific
and Cultural Organization, Place de Fontenoy, 75700 Paris
and Kogan Page, 120 Pentonville Road, London N1 9JN

© Unesco 1981

ISBN 0 85038 447 8 (Kogan Page)
ISBN 92-803-1092-5 (Unesco)

Printed in the United Kingdom by
The Anchor Press, Tiptree, Essex

British Library Cataloguing in Publication Data

Johnstone, James N.
    Indicators of education systems.
    1. Educational planning
    I. Title
    371.2 LB1025

ISBN 0-85038-447-8

Contents

List of Tables ix
List of Diagrams xi
Foreword xiii
Preface xv

Chapter 1: The concept of an indicator of an education system 1
  1.1 Definition of indicators 2
  1.2 Some uses for indicators 6
  1.3 Types of indicators 15
  1.4 Additional advantages and disadvantages of developing indicators 23
  1.5 A framework for defining education system indicators 24
  1.6 Considerations to make when constructing indicators 33

Chapter 2: Statistical issues deriving from education systems data 39
  2.1 The nature of education systems data 40
  2.2 The nature of the scales used for measurement 45
  2.3 Specific features of education systems data 50
  2.4 Internal validity and operational definitions 55
  2.5 Reliability and the consistency of measurement 59
  2.6 External validity and the independence of concepts 69
  2.7 Résumé 70

Chapter 3: The theoretical definition and development of indicators 71
  3.1 Implications of alternative strategies in developing an indicator from a concept 71
  3.2 Norm-referenced indicator formation 78
  3.3 Self-referenced indicator formation 101
  3.4 Criterion-referenced indicator formation 116
  3.5 The problem of interchangeability 118
  3.6 Résumé 121

Chapter 4: The empirical definition and development of indicators 123
4.1 Norm-referenced indicator formation from ratio and interval scales 125
4.2 Norm-referenced indicator formation from ordinal and nominal scales 132
4.3 Résumé 136

Chapter 5: Indicators developed for national education systems 137
5.1 Indicators of national education system inputs 138
5.2 Indicators of national education system processes 149
5.3 Indicators of national education system outputs 167
5.4 Indicators of education system-societal congruence 175
5.5 Technical difficulties in computing indicator values 181
5.6 Résumé 187

Chapter 6: Indicators developed for regional education systems 189
6.1 Characteristics of regional indicators 189
6.2 Indicators of regional education system processes 192
6.3 Indicators of regional education system outputs 198
6.4 Résumé 203

Chapter 7: The presentation of indicators 205
7.1 Methods of presentation involving a single indicator 207
7.2 Methods of presentation involving the inter-relationships among indicators 215
7.3 Methods of presentation involving a number of indicators 218
7.4 Résumé 220

Chapter 8: The use of indicators in educational planning 224
8.1 The use of indicators to help evaluate planned national system development 224
8.2 The use of indicators to monitor regional system development 234
8.3 Résumé 242

Chapter 9: The use of indicators in researching education systems 245
9.1 The selection of indicators to measure societal characteristics 246
9.2 The method of analysis 247
9.3 The short-term relationships 247
9.4 The long-term relationships 253
9.5 Résumé 255

Chapter 10: The use of indicators to classify education systems 257
10.1 Basic approaches to typology formation 258
10.2 Methods of empirically forming typologies from a set of indicators 260
10.3 Cross-national typology formation from input indicators 262
10.4 Intra-national typology formation from process indicators 269
10.5 Résumé 273

Chapter 11: Implications for the future 274

Appendix A: The calculation of values for the Gini coefficient 280

Appendix B: Example of indicator development by the method of factor analysis 285

Appendix C: Technical details of forming typologies 297

Bibliography 302

Names index 311

Subject index 314

List of Tables

1.1 Summary of education variables used in selected comparative education studies 11

1.2 Number of third-level students per 10,000 population – independent South American countries 19

1.3 Statistics describing the percentage of students who are female in Indonesian secondary schools – 1974 22

2.1 Some differences in values of UN and Unesco reported data 53

2.2 Australian first-level enrolments as reported by the Australian government and Unesco 53

3.1 Use of arithmetic mean and geometric mean to measure change in indicator values 75

3.2 Indicator formation by the method of simple addition – female involvement index 79

3.3 Indicator formation by the method of adding standardized variables – female involvement index 83

3.4 Sample of features considered to ascertain the extent of community involvement in an education system 86

3.5 Sample component questions for an indicator measuring the satisfaction of individuals with the education system 86

3.6 Basic statistics for the percentage of six year old children enrolled in the four regions in Thailand – 1960 and 1970 90

3.7 Percentage distribution of first-level students by grade – El Salvador and Nicaragua, 1965 93

3.8 Measures of concentration and percentage enrolments by faculty – selected Asian countries, 1965 and 1971 98

3.9 Basic values and index numbers for indicator measuring human resource involvement in the provision of education – Bulgaria, 1960 to 1972 103

3.10 Education cost and product indicators for Kenya – 1965 to 1972 104

3.11 Indicator measuring change in perception of central educational problems in the USA – 1970 to 1976 108

3.12 Promotion and repetition rates for first-level education – males, Saudi Arabia, 1969-70 and 1973-74 111

3.13 Estimated years of education to be delivered at the first-level – Saudi Arabia, 1969-70 and 1973-74 112

3.14 Summary of possible methods for developing theoretically defined indicators 122

4.1 Orthogonally rotated factor matrix of 1960 input variables 126

4.2 Factor scores for each country on the 1960 input dimensions 128

4.3 Farrell's scalogram of educational structural differentiation: 49 nations, 1960 135

5.1 Ratings of perceived importance of four education goal areas – USA 149

5.2 Summary of the core indicators for the input sub-division of education systems 150

5.3 Rank-order correlations between the structural differentiation indicator and the 1960 empirically defined dimension set 163

5.4 Summary of the core indicators for the process sub-division of education systems 164-65

5.5 Gallup poll ratings of US public schools 1974-76 173
5.6 Summary of the core indicators for the output sub-division of education systems 175
5.7 Social and private rates of return by educational level 180
5.8 Summary of the core indicators for the educational system-societal congruence sub-division of education systems 181
5.9 Example demonstrating the implications of various methods to estimate rates of change 183
6.1 Summary of factor matrix for Indonesian school quality analysis 194
6.2 Summary of variables used for each school and its school district 196
6.3 Variables defining the Barrows and Shaffer education index I, Wisconsin, USA 197
6.4 Variables defining the Barrows and Shaffer education index II, Wisconsin, USA 203
7.1 Academic production index – values and basic data, Australia, 1960-76 215
7.2 Guide to the interpretation of Diagram 7.6 222
8.1 'Involvement of human resources' input indicator – values and basic data, Saudi Arabia, 1965-73 226
8.2 'Availability of student places' process indicator – values and basic data, Saudi Arabia, 1966-73 227
8.3 'Female involvement in the provision of education' process indicator – values and basic data, Saudi Arabia, 1966-73 229
8.4 'Financial involvement' input indicator – values and basic data, Saudi Arabia, 1965-73 229
8.5 'Orientation of second-level studies' process indicator – values and basic data, Saudi Arabia, 1966-73 230
8.6 'Orientation of third-level studies' process indicator – values and basic data, Saudi Arabia, 1966-73 232
8.7 Summary of values for selected process indicators monitoring the development of education systems in selected provinces in Indonesia – 1974 and 1976 236
8.8 Indicator values measuring the equality of the distribution of qualified teachers across Indonesian provinces – 1974 and 1976 239
8.9 Correlation matrix for process indicators – Indonesian provinces, 1976 241
9.1 Summary results of canonical analyses between the Russett indicators and the 1960 and 1969 sets of education system indicators 248
A.1 Sample calculation for the Gini coefficient – data for El Salvador from Table 3.7 – distribution of first-level enrolments by grade 282
B.1 The process variables – their interpretations and codes 287-90
B.2 Summary of factor loadings of variables defining the 1960 process analysis factors 291
B.3 Factor scores for each country on the 1960 process dimensions 292-93
C.1 Values for four input indicators calculated for a sample of ten countries 298
C.2 Matrix of D^2 values for a sample of ten countries 299

List of Diagrams

1.1 A framework for the development of indicators of education systems 26
1.2 Hypothetical profiles for two countries across the five educational areas 32
2.1 Sources contributing to reducing the reliability of education system variables 60
2.2 Population pyramid for Indonesia – 1971 64-5
3.1 Hypothetical inter-relationships between three variables and a concept to be measured 73
3.2 Lorenz curves measuring inequality of distribution of enrolments across grades at the first-level – El Salvador and Nicaragua, 1965 93
3.3 Lorenz curves measuring inequality of distribution of enrolments across grades at the first-level – Nicaragua, 1965 to 1972 115
3.4 Two forms of an indicator measuring involvement in second- and third-level education 120
5.1 Summary results of crude clustering on Harbison and Myers' correlation matrix 139
5.2 Summary results of crude clustering on Harbison *et al* correlation matrices 141
5.3 Summary results of crude clustering on McGranahan *et al* correlation matrix 142
5.4 Comparison between methods for estimating rates of change 183
7.1 Examples of different ways to represent a single indicator 208
7.2 Regional price index by province – Indonesia, 1974 210
7.3 GNP *per capita* by major regions, 1974 212-13
7.4 Comparison of methods for graphing values for an academic production index – data for Australia, 1960-76 214
7.5 Socio-economic profile of Venezuela, 1960 217
7.6 Representation of empirically defined education system characteristics – selected nations, 1960 221
8.1 Lorenz curve for the distribution of qualified teachers at the senior secondary level across provinces – Indonesia, 1976 240
9.1 Inter-relationships between the Russett and the 1960 educational indicators 248
9.2 Canonical vectors relating the Russett 'Economic development' indicator to the 1960 'Extent and refinement of basic education' indicator 249
9.3 Canonical vectors relating the Russett 'Communist influence' indicator to the 1960 indicators measuring the 'Extent and refinement of basic education' and the 'Extent of second-level general education' 250
9.4 Canonical vectors relating the Russett 'Intensive agriculture' and 'Catholic culture' indicators to the 'Extent of female involvement' and 'Development emphasis on quantitatively expanding teacher education' indicators for 1960 252
9.5 Inter-relationships between the Russett and the 1969 educational indicators 254

9.6 Canonical vectors relating the Russett 'Catholic culture' indicator to the 1969 'Extensiveness of educational provision' indicator 255
10.1 Unadjusted cross-national typology across input indicators, 1969 264
10.2 Final cross-national typology based on input dimensions, 1969 266
10.3 Unadjusted and adjusted typologies across process indicators – Indonesian provinces, 1974 270
10.4 Profiles for the three identified clusters of Indonesian provinces, 1974 272

Foreword

There is a growing interest among educational planners and administrators today in the topic of development indicators, and an increasing awareness of how they can potentially be used. In this book, James N Johnstone presents a comprehensive review of the literature on indicators of social development, with a particular emphasis on their application to education systems. He takes into account not only the many books and articles published on the subject in recent years, but also the main results of a research programme carried out by Unesco from 1969 to 1976 on 'A general methodology of human resources indicators', which resulted in several seminars and papers not easily available to social scientists interested in the subject.

This book provides a valuable contribution to the literature on indicators on many counts. First, the author makes a clear distinction between variables and indicators, two terms often confused in other texts, or at least not explicitly formulated. Equally important is his discussion of the validity and reliability of indicators, from the statistical point of view. He also introduces useful notions on the theoretical basis and the development of indicators, such as norm-referenced, self-referenced and criterion-referenced indicators.

For the development of global indicators for the educational system, he uses the systems approach, distinguishing inputs, outputs and processes, and this classification is applied to the indicators of educational systems at the regional level, with an interesting effort to establish the links between global and regional indicators. In Chapter 7 will be found a useful discussion of ways of presenting indicators, which is sometimes absent in other works on the subject, and finally the author concludes with a presentation on the uses of indicators for educational planning purposes, for research on educational systems and for their classification and typology construction. Based on a thorough review of the existing literature on development indicators, the discussion is illustrated by many examples dealing with developing countries.

It is hoped that this book will be of interest to social scientists dealing with development issues, as well as to planners who need to take advantage of the increasing knowledge of synthetic measures of social and economic development, and to adapt it to their specific needs.

Michel Debeauvais
Paris, January 1981 Director, IIEP

Preface

The last two decades have seen education systems in all countries of the world change dramatically. Problems have arisen in some countries which are similar to those found elsewhere. Other problems however are unique to a particular country because of the political, social, administrative or even geographical characteristics of that country. In order to solve their problems, many countries have established divisions of planning. Some of these divisions operate at the national level while others are at the provincial or regional levels. Some (unfortunately not all) of these divisions seriously attempt to plan for the short and for the long term future within their jurisdiction. They do this in the educational sector partly to ensure more people – whether they be children, adolescents or adults – receive a proper and appropriate education, and partly to ensure that education is delivered in the most efficient and effective way.

Within such planning divisions as these, one problem which has been realized -- although it is a problem which is present in all education systems – is the difficulty in measuring education systems and in assessing their development and change. Part of this difficulty can be attributed to people with different orientations having their own perspectives on what should be measured. Politicians, administrators and researchers use different criteria to assess education systems. Certainly some differences must be allowed because the objectives of these assessments are not necessarily all the same.

The complete lack of agreement as to how education system characteristics should be measured is now however hindering proper discussion about the future provision of education systems. People can no longer continually maintain their own amendments independently of others. In many countries, budget allocations for the education sector are no longer rising in real expenditure terms. This constraint which is being imposed by governments is a realistic one in many situations and it can be expected to continue. It is however compounding the problems of planners who operate in properly functioning divisions of planning. The traditional approaches to the planning and the evaluating of education systems have relied on the amassing of large volumes of unco-ordinated and/or unrelated data on particular aspects of systems. Reporting has then been based on traditionally selected and acceptable features from these data banks. Despite the amount of data generally collected, however, it is difficult to answer basic questions such as the following for most education systems:

- ☐ Who is entering the education systems?
- ☐ What are the hopes and aspirations of the entrants and of the society?
- ☐ What equality of access do students have to qualified teachers?
- ☐ Where do the funds for education come from?
- ☐ How much education is produced annually by a system?
- ☐ Are people satisfied with the education being offered?

If answers to questions such as these were to be forthcoming for a particular system, they would generally be like the data – unco-ordinated, unrelated and very specific.

With the changed circumstances in which education systems now find themselves, the traditional collection and reporting procedures must be changed so that data which are collected communicate meaningfully and succinctly on what is happening within an education system. Monitoring of answers to questions such as the above over a long time period informs all people concerned with education systems about how a system is developing and whether policies and plans formulated to guide the future development of systems are really having the impact it was considered they would have.

The communicating of information about education systems in meaningful and succinct ways as recommended can only be achieved through identifying those general or global characteristics which really capture the essence of a system's operational characteristics. If this identification is possible – and it is the thesis of this book that it is – then discussion of the overall system features can be undertaken without the necessary specific qualifications to every comment by referring to particular levels or particular regions or particular sorts of people, etc.

Indicators provide the necessary methodological framework within which the development of meaningful comment can proceed. They are more global than individual variables and hence they avoid specificity of comment. In no way, however, do they confuse characteristics which should be examined in a critical evaluation – at least not if the indicators defined are properly constructed. Nor are they vague in the comment they make.

In the following chapters, the principal objective is to explore the major considerations which arise in conjunction with the term 'indicators'. To fulfil this objective, the term is defined and some ways in which indicators can be developed are identified and the implications of these are discussed. Sets of indicators which may be suitable for national or regional education systems are then outlined. Finally, ways of presenting and using indicators are demonstrated.

The audience for whom this book has been written consists of two distinct groups of people. On the one hand, the book is aimed at educational planners in developing and developed countries alike. On the other, it is aimed at researchers in the fields of educational planning and comparative education. Both of these diverse groups should be concerned with the development and use of education system indicators, for both have the need to measure and evaluate education system characteristics.

In developing indicators, it is essential to use some mathematics and statistics and to have a conceptual understanding at least of the implications of actions taken. One aim in presenting the ways in which indicator development might take place, therefore, has been to outline these conceptual implications yet, at the same time, reduce the complexity of the mathematical description. Generally the computations involved are very simple. It has, however, been necessary to include a few more complex statistical methods to ensure a degree of completeness in presentation.

The initial impetus for this book came from two former staff members of the International Institute for Educational Planning. For their part in urging me to undertake this project I would like to thank Hans Weiler and Neville Postlethwaite. In addition I would like to express my appreciation to Hugh Philp, Mike Bailey (both of Macquarie University), B Fredriksen (of Unesco) and M M Hornibrook for reading selected parts of the manuscript; and to Debbie O'Mara and Barbara Barker for collecting some of the data used. Finally, my greatest appreciation must necessarily go to my wife Marea. She has helped greatly in this and in my other undertakings in so many tangible and intangible ways. For her patience, understanding and support I owe my greatest gratitude.

JNJ

Chapter 1

The Concept of an Indicator of an Education System

As a consequence of societies becoming increasingly more complex, there has been a tendency for the main characteristics of each of the sub-systems of societies to be less clearly identifiable. In addition, the relationships among the sub-systems have also become less and less clear. Previously, for example, it was possible to describe an education system and the broad relationships between education and the economy of a country and between education and certain societal characteristics with some degree of precision. As development has proceeded however, these characteristics and their attendant relationships have become less easy to identify. Things are now very much more complicated and there is a multitude of factors within each societal sub-system. Each of these intra-system factors exert forces on the other factors in the same system as well as on factors in the other sub-systems. One consequence of this increase in complexity is that the inter-related tasks of planning and policy formation both within and across the various sub-systems of society are now made very much more intricate.

A central difficulty which must now be addressed in planning the various sub-systems is that of ascertaining whether or not development in the various sub-systems has occurred. Is, for example, the education system accommodating the demand for education by individuals in a society? Has the ability of the population to communicate increased or decreased? Have housing conditions improved? Have the rates of incidence of various diseases changed? Have prices increased? Questions such as these need to be answered so that plans and policies can be formulated to facilitate development towards the achievement of certain agreed societal goals.

Answers to questions such as those above must be provided in special ways. It is not desirable, for example, to provide answers involving massive amounts of specific detail. In any situation, very many aspects of each facet of any sub-system can be identified. In relation to the question on communication ability, what aspects should be reported? Some important considerations are the ability in literacy and numeracy of the population, the provision of various kinds of telegraphic facilities, the availability of books and newspapers and the opportunity for free public debate and discussion.

To overcome the problem of having to analyse massive amounts of information in order to obtain general or overall answers to questions concerning development, the concept of an indicator has emerged. Although this concept has been commonly referred to for many years, it is only in recent times that its proper connotations have been developed and more fully acknowledged.

1.1 Definition of indicators

The issue of defining what an indicator really is has been debated at great length in the literature. Land (1975) takes up some of the points in this debate and demonstrates the conflicts between the more commonly accepted definitions. No such undertaking is attempted here, however. Instead, a more conventional, although perhaps less academically elegant, approach is followed in which the term 'indicator' is employed in accordance with its everyday usage.

The Oxford dictionary defines an indicator as 'that which points out or directs attention to something'. Webster's dictionary defines the word as 'something which points out or points to with more or less exactness'. These definitions are similar and both reflect the essential nature of what, in the social sciences, should be properly termed an indicator.

Indicators make general comment

One feature to note from these definitions is that an indicator does not necessarily give a precise comment on the state or nature of the thing being examined. Instead, a much more general impression is conveyed 'with more or less exactness'. This is similar to the everyday situation in which directions are requested on how to get to a particular place or address. Here indications are given such as 'the office is down the street about 500 meters or so' or 'take the first to the left, second to the right and then straight ahead – you can't miss it' or 'over there'. There is nothing really exact, in the scientific sense, in any of these directions. They do however reflect the way in which an objective can be achieved as well as to what degree approximately the objective has been achieved at any intermediate stage. If a more exact measurement of progress is needed, this can be easily obtained. For the purpose of arriving at a given destination, however, such general directions (or indications) are quite sufficient.

Concepts analogous to that implied for social science indicators exist in other fields of science. In chemistry, litmus paper is used as an indicator of the acidity or alkalinity of a liquid. When red litmus paper is placed in an alkaline solution, it turns blue; when blue litmus paper is placed in an acid it turns red. Hence by placing the two types of paper in an unknown liquid and then noting what colour change occurs, a conclusion can be drawn about one important property of the liquid. Normally this overall indication is sufficient. A more precise estimate of the acidity of the liquid can, however, be made by chemically testing for what is termed the 'ph value'. This estimate basically involves looking at the reasons why the litmus paper took on its blue or red colour.

The above analogy helps to elucidate what should be the nature of an indicator in the social sciences. It should be something giving a broad indication of the state of the situation being investigated. It does not necessarily do this with a high degree of scientific exactness but it does indicate the general state of what is being examined. If very particular information is required on some component, a re-analysis of the situation directed explicitly to determining only that particular information can be undertaken.

An example demonstrates the point being made. One of the indicators for a national education system discussed in a subsequent chapter is 'the involvement of human resources in the provision of education' (see Table 5.2). This indicator measures the participation rates of pupils for all three levels of an education system as well as the extent of recruitment of teachers from the available work-force. A value calculated for this indicator must be interpreted globally as reflecting the overall expansiveness of involvement of human resources in an education system. It cannot be interpreted as commenting on any one level of education. If specific information is required on one aspect of the total involvement – for example on the net first-level enrolment ratio for females – then this can be calculated and interpreted separately.

Indicators are distinct from variables

Definition of an indicator as being a global quantity in this sense is in direct opposition to many of the ways in which the term is presently used. Some statistical reports, for example, confuse the term 'indicator' with the term 'variable'. In these, data are reported separately for such phenomena as:

- ☐ population aged 15 years and over studying for post-school qualifications
- ☐ pupil-teacher ratios
- ☐ proportion of labour force who hold a second job
- ☐ number of aged pensioners by age and sex
- ☐ aggregate family income by size.

All of these variables are particular facets of society. Although they are important facets, no conclusion can be validly drawn on the state of the education, work-force, social security or economic systems from any one of them. They are variables and they thus particularize. In contrast, an indicator would combine one of these variables with other conceptually related variables to form a general overview of the broad features of that aspect of the system being described. This distinction between indicators and variables is crucial to the ensuing discussion.

Definition of indicators in this way does not require that a single indicator be calculated for each of the main societal sub-systems. Indeed there may be 20-30 indicators for education, for health, for social security, etc. There are not, however, hundreds of such indicators for each sub-system as there would have to be if individual variables were defined. Indicators therefore present a summarized picture of what a system is like so that a general overview is gained rather than a very specific and fragmented series of comments.[1]

Wilbur J Cohen, after reporting two experiences which influenced his attitudes to presenting data in a form whereby they could be effectively used for policy decisions, concludes that simple answers should be given to simple questions. If 'too much statistical information [is given] at one time, it casts serious doubts on your competence as a research person' (1975, 58). He argues that for policy purposes, exact and complete detail qualified by various measurement constraints is generally not needed. Instead, a general impression conveyed 'with more or less exactness' is to be preferred. This then is precisely the role an indicator should be designed to fulfil.

An indicator value expresses a quantity

A third feature of an indicator is that it is something which is quantifiable. It is not a statement describing the state of a system. Instead it must be a real number to be interpreted according to the rules governing its formation. Hence the number '96' may indicate a relatively low level of some quantity being measured. For another indicator, the value '2.8' might imply a very high level of development. The rules governing the construction of indicators and the number system mapped by the chosen indicator must thus be very clearly and explicitly stated.

Indicator values are temporal

A fourth feature of an indicator is that any one of its values applies to only one point or period in time. For some indicators, this point may be a particular month or a given year. For others, the value may relate to development occurring over a three month or a five year period. For full interpretation of indicators, a large number of observations should be available. Generally these observations will relate to the same system. Hence some indicators may have a series of values calculated for them covering ten or more years. These time series allow monitoring of the evolution of the system being investigated. Both positive and negative changes in the system can then be detected, and periods during which no change has occurred can be identified. In other situations, data for a number of similar systems might need to be collected. The interpretation of an indicator value is then facilitated by referring to the values pertaining to these other systems for the corresponding time period.

---

[1] This distinction between variables and indicators is of utmost importance for describing the characteristics, construction approaches and the use of indicators in planning and research contexts. Very often, however, the behavioural properties of the two types are very similar. Hence as indicators of education systems, examples will often be discussed on the following pages in terms of variables rather than in terms of indicators.

Indicators are basic units in theory development

Another way in which to view an indicator is from the point of view of research work directed towards theory building. In such work the steps often characterizing the first stages are:

1. A statement or a tentative description of the phenomenon to be investigated;
2. The identification of the central concepts of relevance to the particular phenomenon;
3. The operational definition of the concepts into measurable variables;
4. The planning of data collection and indicator construction.

In stage 2, the central concepts identified are generally very global in nature. They reflect a broad characteristic involved in the phenomenon rather than a very specifically defined characteristic or variable. Terms used at this stage in research include such broad notions as intelligence, self-concept, the economy, modernization and centralization. What then occurs in stage 3 is that these broad concepts are defined as being represented by a number of particular variables. Intelligence may be represented by a transformed score obtained from one or more sets of test items; modernization may be represented by the time when political power transfers from traditional to modern leaders in a society (as is done by Black, 1967). These variables are then used in subsequent stages of the research endeavour and generalizations are made from the relationships established among these variables to the relationships which may exist among the broader concepts.

Unfortunately what often happens is that imperfections occur in the operational definition of a concept. The meaning implied in a concept additional to that conveyed by a variable is then lost in the definition used for data collection although this loss is ignored. For example, when a positive correlation is found between the first-level enrolment ratio and gross national product per capita across a number of countries, many researchers are tempted to conclude something about a relationship between educational provision and economic growth or development. This conclusion, although tempting, is fallacious. All that has been demonstrated is a relationship between two variables.

In the research context therefore, a real danger exists in substituting a variable for what is, in reality, an indicator. As an indicator – rather than a single variable – more properly represents the concept, it should be used where possible as the elementary unit to establish relationships between domains or sub-systems. This is not to say that only indicators contributing towards theory building should be developed. Instead it is to note that indicators fitting into a theoretical context can be of major importance. In order to identify these indicators, considering the present low level of theory development, it is essential to construct a variety of properly developed indicators and then determine both their real and their contextual meanings. This has to involve cross-temporal and cross-system analyses so that differences between general and specific sources of variation can be identified.

Dubin (1969, 44-6) refers to an indicator in the sense implied here as being a basic unit in theory building. He considers that there are a number of stages in the development of such units. Originally, for example, they can be termed 'primitive units' because they form the starting points from which a theory might be developed. Later, after proper investigation and development, they can be termed 'sophisticated units' because they are more fully and adequately defined. In this transition, and also beyond it, relativities between units must be hypothesized and tested. Proper generalizations from results can then be drawn.

Perhaps because of the close relationship between theory and indicators, a comprehensive theory of education systems development has not yet begun to emerge. This lack reflects Coleman's observation that 'quantitative laws and theories seem to require more in the way of measurement than social phenomena are prepared to give' (1964, 55). A basic assumption made here to support the development of indicators is that education systems can and must be measured in adequate, reliable and valid ways. Indicators so developed can then in turn be used to formulate quantitative theories.

1.2 Some uses for indicators

Indicators can be constructed and used for a wide variety of purposes. Many indicators can be used by planners, administrators and researchers in very similar ways and for very similar purposes. This does not always have to be the case at all times, however. Planners, for example, may require particular kinds of indicators from either administrators or researchers on particular occasions. When these different purposes for using indicators exist, they must be made explicit and given due recognition. If they are not, the development of indicator sets for education systems will be such as to fulfil the purposes of one group but not the others. This will in turn lessen the potential contribution indicators can make.

The statement of policy

One use for indicators is the statement and development of policies in a more definite and coherent way. One of the major problems in planning – and one which is manifest in many plans – is the lack of specificity. Policies and their objectives are often stated in such general and vague terms that it is unclear precisely what a plan is designed to achieve. Examples of such statements are:

- ☐ to help all persons reach their fullest potential as unique individuals by meeting their educational needs
- ☐ to develop further courses in science and technology
- ☐ to provide equality of opportunity to all the nation's citizens
- ☐ to cultivate amongst students in the first-level, harmonious physical and mental development.

None of these statements indicates precisely what change is to occur within the system. In any evaluation either of a plan or of the development of a system – and such evaluation is regarded here as being a *sine qua non* of any plan – an assessment of what the system has attained is rendered impossible. What constitutes an individual's 'fullest potential'? In what ways are science courses 'developed further'? It may be in terms of increased enrolments or changed curricula or different teaching styles or even some other development. Specifications of this kind must be made so that change in the present system can be measured and the attainment of these planned and expected outcomes ascertained.

The lack of specificity in plans for education system development is, unfortunately, widespread. Drewnowski (1970), for example, attempted to analyse certain features of broad planning proposals by examining 61 countries each of which had prepared development plans. The lack of statements attempting to specify plan objectives forced him to reduce his investigation to cover only 14 countries – a very considerable reduction.

Certainly the argument can be sustained that it is not always necessary for development policies and plans to state goals which are to be attained during the period of the current plan. This would be the case when the statement of a goal is really the statement of an intention to do something in the future. Such intentions might be, for example:

- ☐ to promote a co-ordinated system of higher education, including university education, in accordance with the nation's needs and universal standards of excellence
- ☐ to educate the outstanding youth, refine their civic morality and give them cultural, scientific and military training
- ☐ to impart a broad general education to all children.

The statement of intentions such as these is quite proper and is to be encouraged. The policy statements concerning what is to be attained by a particular plan, however, should state precisely what quantity is to be changed and what level that quantity *is* to reach by the end of the plan. If statements of commitment do not indicate planned outcomes, some of the basic premises of the planning process itself must be questioned. Ferriss sees the integration of goals and indicators as being essential. He states 'serious planning of the education system should always be accompanied by the development or use of indicators measuring the degree to which the goal of a plan is achieved' (1969, 6).

Basically there are two ways in which policies and their objectives can be formulated to enable indicators to be used in the planning process. In one way, policies and objectives can be phrased in general terms although the features of the education system to which the objective is directed are identified specifically. Thus indicators are defined from the general statements of objectives. The defined indicators are then used to assess the extent to which the more generally stated objective has been realized by the end of the development plan.

The other way in which indicators can help to clarify the nature of the intended outcomes of a plan is through their direct incorporation into the stated policies and objectives. This then makes the policy and the attendant objectives more specific in terms of both the features to be changed and the levels to be reached. Seers summarizes his view of this approach by stating that 'the most important use of development indicators is to provide targets for planning' (1972, 32). The targeting approach is the ultimate in specificity of objectives and for this reason it is sometimes considered to be less elegant. In many situations, however, the establishment of targets and the specification of these in the statement of objectives is essential. When resources are limited, such an approach identifies those particular areas which are to receive investment priority.

In whatever form they are used, indicators have a real part to play in planning education system provision. Their potential contribution is that they can force planners, policy makers and politicians to state with more precision what future situations should be like. Without this precision, planning will continue to be as ineffective as it has been so often in the past. Cohen views the potential role that indicators can play as follows: 'accurate meaningful indicators are a prerequisite for valid answers and for planning effective remedies when needed' (1967, 87).

Monitoring education systems

A second use for education system indicators is for the monitoring of changes in systems. This is perhaps the most common purpose envisaged for them. It is highly desirable that changes in features such as enrolment ratios, level of female participation and the levels of comprehension of the national language should be detected as quickly as possible. If changes occur, these can either be publicized to show the favourable development of the system or remedial action can be taken before the problem is aggravated through moving beyond the easily remedied stage.

Coombs insists that such indicators must be developed for the field of education. He sees that education systems do not merely grow but they must change as well. Hence he posits that:

> one of the necessary steps towards identifying the right questions . . . is to develop better instruments for assessing the performance of education systems and their various parts . . . . What sorts of indicator [does] an education system need to give itself an annual check-up? Without a battery of indicators of an appropriate sort, educational planners and policy-makers have little to go on except hunch and prejudice. (1969, 33-4)

The 1969 Report of the National Academy of Science comments in a similar vein by recommending that 'the social indicators adopted should be sufficiently precise to reflect annual changes' (p 105). Others making analogous comments include Baster (1972, 3-5), Cazes (1972, 20) and Culyer *et al* (1972, 94). In a non-educational field, Colley argues that 'a continuously reported social change index for each community . . . could reveal the need for social remedial action before the social conditions in a community become critical' (1975, 94).

Change, as envisaged here, is not being attributed to a particular cause. It could therefore be change as a result of the natural evolution of the education system or change as a result of the implementation of certain policies. Indeed, the use of indicators to monitor change must be seen in a broad perspective. Too often, a very narrow view of educational development – and thus change monitoring – is taken. Some of the ways in which a system develops are planned and anticipated in plans and policies. Accompanying these developments, there are also unplanned but anticipated outcomes as well as unplanned and unanticipated outcomes. In addition, any goals established for a system implicitly assume that progress will be made without any deterioration of the ways in which the system is operating at present.

One situation demonstrating the need for such monitoring is where a country wishes both to expand its education enrolments at the first- and second-levels for males and females and to ensure that a particular balance continues to exist between the two levels. To determine whether this objective is achieved is relatively straightforward. Either several specific variables can be defined and data collected on them during the period covered by the plan or two general indicators can be defined to assess the overall impact of the policy. This approach is really the use of indicators outlined previously.

Implementing this policy, however, could result in a situation in which the objective of increasing enrolments was achieved at the expense of a decline in overall involvement in education (as measured by the various enrolment ratios). Normally an occurrence such as this would not be noticed until an official publication was being prepared and an administrator or planner – or the public at large – drew attention to the attendant and unanticipated outcome.

It could be argued that the situation just described would happen in very few countries. This may be so but it could happen in some countries at least and, for them, the indicator measuring involvement would be a useful adjunct to the normal statistics included in any annual report provided to a minister of education. In countries where both enrolment and involvement properly increased, other unplanned outcomes might occur instead. These could relate to:

- ☐ a change in the emphasis on a particular type of education at either the second- or the third-level
- ☐ a proportionately greater increase in expenditure for education than was anticipated
- ☐ a decrease in pre-school education enrolments due to parents withholding their children because they do not consider this form of education wholly desirable especially when entry to the first-level of the system is made easier.

This does not imply that all contingencies can ever be anticipated and investigated in the context of an annual reporting system. Certainly they will not be when the orientation is towards either reporting only gross numbers or ascertaining only progress made towards stated goals. The assertion is made however that a standardized set of indicators can be defined and used to monitor broad development patterns so that general indications of the ways in which a system is evolving can be obtained on a regular and consistent basis.

An awareness of the need to develop indicators for monitoring education systems is beginning to develop. In the United States, for example, the US Congress has amended the Elementary and Secondary Education Act to require an annual report from the assistant secretary for education. This report is to comment on 'the condition of education in the United States'. It is therefore a report which will monitor educational development patterns. Although many researchers have in the past demanded indicators be developed for such purposes (Turnbull, 1975, is one of these) it is to be hoped that more countries will soon follow the lead of the US in legislating for the monitoring use of educational indicators.

Researching education systems development

A third purpose for developing and using indicators is for representing education system characteristics in research studies. Two broad types of study have had substantial problems in the past in describing the development level and nature of education systems. One kind of study involves the cross-national analysis either of educational patterns or of broad inter-relationships between societal sub-systems, one of which is education. In these studies, education systems are represented by any one of a number of variables. Table 1.1 presents a selection of variables

Education – economy studies		**Education – socio-political studies**	
Study	**Education variable**	**Study**	**Education variable**
Edding (1962)	% of GNP on education	Jahoda (1961)	level of education reached
Svennilson *et al* (1962)	enrolment ratio	Fox and Miller (1965)	combined first- and second-level enrolment ratios
Bowman and Anderson (1963)	% of population in primary education % of population in post-primary education	Almond and Verba (1963)	years of schooling
Harbison and Myers (1964)	combination of second- and third-level enrolment ratios		
McClelland (1966)	secondary school enrolment per 10,000 population	Cutright (1963)	combination of literacy rates and higher education enrolments per 10,000 population
Meyer (1971)	ratio of enrolments at each level	Rummel (1972)	expenditure on education as a % of government budget first-level pupil-teacher ratio first-level unadjusted enrolment ratio second- and third-level students per 10,000 population

Table 1.1: *Summary of education variables used in selected comparative education studies*

used in research studies investigating the relationship between the education and the economic, political or social sub-systems. The eleven studies listed use fifteen different variables for the purpose of representing the education system. As the correlation coefficients between pairs of these variables are all considerably less than 1.0, the use of one variable rather than another could lead to differences in the results obtained. For example, if Meyer had used McClelland's or Cutright's variables to relate to his measure of a country's economy – gross national product per capita – his results would have been different.

The other type of research study which has difficulty in representing education system characteristics is that directed towards measuring change and development patterns within individual countries. Although this area does not have to be the prerogative of researchers, it has tended to be so. Some sort of 'P methodology' approach can generally be adopted here (see Gibb, 1956, or Rummel, 1970b, 198). The advantage of using indicators rather than a series of variables in such a research study is that broad patterns over inter-related variables can be identified. The more usual practice of establishing trends on specific variables can be difficult to justify because the unique importance of variables sometimes cannot be definitely assessed. Not only is it difficult to support the selection of a group of specific variables for individual analysis but each variable is likely to have been differentially affected by the implemention of particular policies. The amount of generalization allowable from the specific trends to the broad general trends is then reduced.

Despite a considerable research effort directed towards identifying causes and patterns of education system development, very little is really known about the evolution of these systems. Contradictory findings have been produced and these may, in part at least, be attributable to studies adopting a univariate rather than a multivariate approach to their analyses. Such an approach has been acknowledged as hindering the development of theories in other fields also. Glock and Stark (1965, 21) claim that the study of religion and the assessment of religiosity has been hindered by researchers looking only at very particular characteristics of individuals and their religion rather than at the major dimensions which define religiosity. Similarly, Tatsuoka and Tiedeman argue that no theory to explain the pattern of interactions within classrooms can emerge as long as researchers continue to study correlations between various pairs of observed variables. The latter writers claim researchers 'tacitly hope that, with a sufficient accumulation of . . . empirical findings, a theory . . . will eventually emerge. According to modern views in the philosophy of science, however, such hopes are foredoomed: theory-building is a deliberate enterprise' (1963, 142). These comments apply equally to the study and researching of education systems. Attention has to date been focused too much on the inter-relationships between particular attributes of systems and not on the multivariate inter-relationships or on the relationships among broad concepts.

One of the implications of not understanding the mechanics of education system development is that planners in developing countries are placed in a dilemma about the best way to develop their plans. Should they, for example, use a targeting approach to plan their provision or should they attempt to implement policies to change the existing balance across different sectors both within and outside the education system? If the former approach is adopted, a developing country will identify another country already at a higher level of development. It will then establish development targets for a medium term period according to the levels prevailing in that more advanced country. For example, Tanzania may select, for a variety of reasons, a country such as India to be its target. Tanzania will thus attempt to reach, by the end of one or more plans, the type and extent of educational provision India has at present. Puerto Rico might choose the United States especially in its structure and balance between levels.

There are many other ways in which research studies should use indicators rather than variables as the basic units for analysis. To date, however, research into education systems has failed to produce more than a few significant results to guide planning decisions.

Classifying education systems

A fourth purpose providing an impetus for the development and use of indicators is to facilitate the formation of a valid and reliable classification of education systems. Such classification is a fairly common undertaking and it can occur for a number of reasons. Politicians, administrators and planners of many developing countries, for example, often find it most informative and helpful to discuss development problems and ways to alleviate them with their colleagues from other countries at about the same level of development. Within the Unesco framework, organizations such as the Twenty-five Least Developed Countries have been established to facilitate mutual co-operation between certain nations. In the Karachi plan, the Asian countries are divided into three different groups so that the same demands are not made nor the same expectations held for all countries participating in that plan. Such distinctions properly admit development reality and are to be encouraged so long as the basis for the grouping of systems is valid in terms of education system characteristics. Agencies such as Unesco, UNDP, World Bank and the Ford Foundation also find it useful to classify countries on the basis of the development levels and/or their type of educational provision. Classifications enable these agencies to make a more objective assessment of the current level of development of the various country systems and to assess the results of past aid programmes.

Classifications can also be made, for similar reasons, of provinces or states within a single country. Such classification can help to demonstrate the extent of similarities and differences in educational provision as measured by a particular set of indicators. Classification of provinces into similar educational levels then leads to an estimate of the nature of any disparities in provision. When traced longitudinally, such a schema can reflect those regions where progress is being made and those areas which are falling even further behind the national average. Such classificatory exercises can also demonstrate the extent to which a national pattern of education exists or whether provincial patterns are so diverse as to make a national pattern indefinable.

Classifications conducted either cross-nationally or intra-nationally require that broad patterns of educational provision form the basis for typology formation. Using a large number of single variables to measure the patterns leads to confusion as to what similarities are really represented in a particular cluster of systems. Classification based on only selected individual variables does not afford the necessary valid basis. Errors associated with single variables can disproportionately influence the composition of any typology. Indicators, when properly formed, minimize the effects of mis-classification due to both of these sorts of error. Greater reliance can then be placed on a typology formed by clustering profiles of education systems across a set of standardized indicators.

Absence of normative comment

Another purpose or role often assigned to indicators is that they should comment on the degree to which a particular goal is being achieved. That is, they should be normative in nature. This purpose is denied here. Indicators, for whatever reason they are developed, should be neutral in the comment they make on a system. It is for the interpreter to make a normative judgement on the indicator value. To cite a relatively obvious example, an indicator should not be developed to comment on the degree to which an education system is becoming decentralized. It should instead comment on the place a system occupies on a scale extending from 'complete central control' to 'no central control'. The scale value for a particular system can then be interpreted in light of stated system aims, current policies, etc. Unfortunately, many indicators can be defined which do proffer normative comment. In politics, some of the 'extent of democracy' indicators are of this kind. The way in which these are defined prevents their use in communist and socialist countries as well as in countries under dictatorships. Such indicators are based on the assumption that democracy is 'a must'. It should instead be admitted by those responsible for defining indicators that many political systems are not striving towards establishing such a political state.

Defining an indicator in a neutral way, as is advocated here, certainly does introduce the possibility of interpreting its value in numerous ways. Any multi-dimensional object, however, can be so interpreted — and consequently sometimes misinterpreted. As Lazarsfeld (1959, 41) points out, even a forest of trees can be viewed in a number of ways. The farmer perceives it as a source of lumber, the hunter as an animal's abode and the nature lover as a source of enjoyment and a place of beauty.

Although certain distinctions are made here to discuss individually the several purposes for developing indicators, the purposes do in fact overlap quite considerably. In practice it can sometimes be difficult to identify whether an indicator is used primarily for planning, monitoring or research. Indeed, it does not really matter. Indicators should be regarded as potentially serving all three types of endeavour. The type of indicator and the method of construction do not vary according to the purpose. Which particular indicators are used, and how they are used, certainly do vary, of course, according to the purpose of the particular investigation being undertaken and the particular questions which are to be answered. On this basis, therefore, the distinction Baster (1972, 11) requires between indicators for research and indicators for planning is not considered valid.

1.3 Types of indicators

Indicators can be classified with respect to many different criteria. Some of these criteria reveal important distinctions in the ways indicators are used or could be used. Others identify concepts underlying the different approaches to indicator construction. It is appropriate at this point, therefore, to review some of these criteria as they have been proposed in the literature. As it is impracticable to include all schemes suggested, only those having a direct relevance for the later methodological issues are outlined.

The first distinction between types of indicators concerns the extent to which component variables contribute to the formation of an index. Three main forms of indicator can be identified: the representative index; the disaggregated index; and the composite index.

Representative indicators

The representative indicator is the most common form of indicator presently used for research, administrative and planning purposes. It involves the selection of a single variable to reflect some aspect of an education system. Amongst the most frequently chosen educational variables are the enrolment ratio for the first-level (or the first- and second-level combined) and the percentage of the gross national product spent on education. Level of education reached and literacy rates are also often used. To date, the selection of a single variable to represent educational level or effort has been fairly arbitrary. No realistic justification for selecting one variable rather than another is generally advanced in research reports or planning documents. Such omission has had the effect of severely limiting the generalizations which can be drawn from research studies as well as preventing proper comparisons being made between countries. Schnore, in his study of modernization, observes that 'prior research has tended to fragment the question [of indicators], for each investigator appears to have been preoccupied with one variable and to have designed his analysis to reveal the interrelations of one factor with all others . . . . Almost no effort has been directed toward simultaneously ascertaining the manifold associations among a large number of indicators of modernisation' (1961, 231-2). Edwards (1970, 251) draws the analogy that the use of one variable or index to represent education is akin to the practice of the early twentieth century psychologists who used a single IQ rating to describe the complex mental abilities of an individual.

To choose just one variable to act as an indicator for an education system must be an impossible task. An education system is an exceptionally complex entity consisting of many different facets. Indeed, as is shown in Section 1.5, many of these facets themselves are even conceptually independent. Hence the use of representative indicators must be considered as a most unsatisfactory practice for planners, administrators, policy-makers and researchers.

Disaggregative indicators

The second form of education system indicator is the disaggregated index. Instead of selecting only one variable to represent a concept, this type of indicator requires the definition of variables for every element or component of the education system space. Ideally, every defined variable should be independent of every other variable so that information is not duplicated throughout the indicator set. The final set should also exhaust the definition of the space. Indicators (again really only variables) classified as forming a disaggregative set, individually give very precise information about each element of a system. The result of completing such an exercise, however, is a very long, complex and confusing list of variables which is impossible to use effectively or efficiently. Some compromise to deal with only one part of the education system space raises questions of insufficiency which are difficult to answer. Additionally, it is almost certain that a parsimonious set would not be chosen and that information, especially that measuring some of the concepts alleged to be more important, would be duplicated many times. As indicators chosen in this way do not allow proper description of an education system as is required by the previous section, this type of indicator is not discussed further.

Composite indicators

The third form of indicator – and the one recommended as being the only justifiable type – is the composite indicator. This form of indicator combines, in some way, a number of educational variables. The final composite is thus interpreted as a 'kind of average' of all variables entering into the combination.

Definition and selection of a composite indicator to represent aspects of education systems admits to the complexity of such systems. There are so many separate yet allied aspects to each of the major concepts of a system that a multivariate description provides the only way in which a manageable description can be realized. Indeed Ferriss goes further than this by claiming that the formation of a composite indicator produces something 'which is more general than the detail composing it' (1969, 12). Ferriss' assertion implicitly acknowledges the fallacy of placing too much reliance on the information conveyed by a single variable. A number of related variables, however, can together describe a feature with a reasonable degree of exactness.

Retzlaff (1965) agrees with this view by claiming that a composite indicator must have greater reliability than a single variable. If this is so, he claims, a composite indicator 'would possibly allow for greater comparability through the process of interchangeability or substitutability of components of the index . . . . This has particular relevance for the problem of missing data as well as the fact that different countries use several different techniques for measuring data of a comparable character' (1965, 68).

A composite indicator thus combines a number of variables to produce a single value for a particular system. It does not redefine the scope of the characteristic being measured in such a way that a different sort of calculation is introduced. For example, an indicator of the involvement of females in an education system might be defined to consist of the percentage of female students at each level of the system together with the percentage of female teachers at each level. Six variables are therefore required. Combining all students and all teachers into a single group and then calculating a gross percentage from this total does not constitute the formation of a composite index. The six variables listed above must be preserved separately because they should comment individually and, to some extent, independently, on the nature of female involvement. If it is considered appropriate, some of the component variables may even be weighted more heavily. Hence a composite indicator amalgamates in some way the contributions from a number of discrete component variables.

Composite indicators possess the best features of both the single variable and the disaggregation approaches to indicator formation. As relatively few composite indicators are needed to define an education system space, the benefits of simplicity of description and relative ease of comprehension, which make the representative approach so attractive, are retained. On the other hand, by combining variables, particularly those that are conceptually alike, information from a broad span of the space is conveyed. This is the same as with the disaggregative indicator but it is much neater and more economical.

Bases for interpreting indicator values

A second scheme for the classification of indicators concerns the basis used for interpreting the indicator value estimated for a particular education system. There are three such bases:

☐ other education systems at the same point in time — and hence a norm-referenced (or synchronic) comparison[1]

---

[1] The term 'norm' is used to imply a comparison with another individual system or with the average standard of a number of education systems. It should not

☐ the same system but at a different point in time – and hence a self-referenced (or diachronic) comparison

☐ an ideal system or a planned system objective – and hence a criterion-referenced comparison.

These three bases parallel different approaches within educational testing practice. There both the norm-referenced and criteron-referenced bases are the ones most used.

Two variables demonstrate the differences in interpretation implied when the concepts of norm- and criterion-referencing are applied to education systems. The variable 'net first-level enrolment ratio' has a maximum value of 100 and a minimum value of 0. To know that Brazil, in 1974, had an adjusted first-level enrolment ratio of about 75 is to imply that about three-quarters of the relative age group (7–14 years) were enrolled in first-level education. Progress can therefore still be made towards providing resources to enable this value to rise in the future so that it is closer to the ideal of 100. Such an interpretation is thus a criterion-referenced interpretation. It is made in comparison to some stated criterion and requires no information related either to other countries or to the value of the ratio for Brazil in other years.

The variable 'number of third-level students per 10,000 population' does not really have a clearly defined maximum or minimum value. Hence to determine that, in 1974, Brazil had a value for that variable of 92 means very little by itself. This may be a high figure or it may be low; it may reveal a good situation or it may reveal the reverse. In isolation no value judgement or interpretation of the value is possible.

One basis for interpreting the value would involve ascertaining a target value Brazil may have established for 1974. This would, as in the previous example, be a form of criterion-referenced interpretation. The value of 92 would be compared to the criterion previously established.[1]

A norm-referenced interpretation is also possible. It requires the estimation of values for the same variable for other countries for the same year (or about the same year). If, therefore, a set of values such as that in Table 1.2 is provided, the value for Brazil can be interpreted relative to the other independent South American countries. In this group, Brazil is just below the average (mean) value of 100 and it is ranked sixth out of the twelve countries listed. A relative judgement is all that is made. Naturally also, if a different group of countries had been chosen as the norm-referenced group, a different interpretation of Brazil's standing would have to be made.

Finally the basis for interpretation referred to as self-referenced

---

therefore be taken in the sense of its other meaning of 'a standard'. This latter interpretation covers the sense reserved for criterion-referenced comparison.

[1] Every country in the world has students at the third-level at least studying in another country. The variable could therefore be defined to take this into account.

Country	Year	Value
Argentina	1975	147
Bolivia	1975	88
Brazil	1965	19
	1970	46
	1974	92
Chile	1975	146
Columbia	1975	63
Ecuador	1971	69
Guyana	1973	30
Paraguay	1972	45
Peru	1975	125
Surinam	1965	20
Uruguay	1975	106
Venezuela	1975	178
		Mean value 100

Source: Calculated from Unesco *Statistical Yearbook*, 1976

Table 1.2: *Number of third-level students per 10,000 population – independent South American countries*

implies that an education system indicator or variable measured on a particular system in or for a particular time period is compared to the corresponding value derived from that same system in or for another time period. Table 1.2 also shows the 1965 and 1970 values for Brazil. Comparing all three values for Brazil shows that a very large increase has occurred over the 1965-74 period. Indeed the value of the variable has risen about 484 per cent. This interpretation of the increase is a self-referenced interpretation. Although it appears to be significant – and it does most certainly reflect a significant expansion of third-level education in Brazil – the value should be compared in a norm-referenced way before conclusions about Brazil's relative expansion are drawn.

No one basis for interpreting indicator values can be recommended. Any one of the three are appropriate depending on the interpretation desired for a particular indicator. For certain planning purposes within a single system – especially those concerned with evaluating plan implementation – either the self- or the criterion-referenced bases might be the most useful. For purposes of monitoring educational development, interpretation in a self-referenced way is best. For cross-national research purposes, norm-referencing often provides the most appropriate basis. The major implication of the distinctions among the three bases is that indicators should be formed by methods best allowing the interpretation desired. Although estimating an indicator value on one basis can sometimes allow other interpretations, a clearer and more direct interpretation is generally obtained if the indicator is formed on the appropriate conceptual basis.

Absolute or ratio measurement

A third distinction useful for identifying types of education system indicators contrasts indicators measuring absolute size and indicators measuring relative size. Examples of variables of the former type are total enrolments, number of school buildings, cash flows, etc. Examples of variables of the latter type are enrolment ratios, variables based on percentages, growth rates, etc.

One disadvantage of absolute quantities is that one system can appear to be demonstrating a greater educational effort than another simply because it has a greater potential to do so. It does not actually have to be exerting a greater effort. For example, a country with 100 pupils enrolled from a total of 1000 potential students who are eligible to enrol is not exerting the same relative effort as a country with 90 pupils enrolled from a total of 100 who are eligible. Admittedly the first country is providing more places in its education system but a comparison based on measures which are independent of potential or comparable baseline values has limited value. What is the point in comparing enrolment totals in India with those in Bhutan or the Maldive Islands?

Ratio values, if properly calculated, do allow comparisons of relative effort by apportioning the effect of differences in potential. In the above situation they indicate the former country (with 100 pupils) has an enrolment ratio of 10 while the latter country (with 90 pupils) has an enrolment ratio of 90. These ratio values do allow a comparison to be made between the two countries which has a very significant meaning.

Indicators and variables using a ratio basis for their definition should be selected and used when it is appropriate to do so.[1] The effects of potential to provide are then removed. This is especially so in a norm-referenced situation because then large systems and small systems can be compared in a more meaningful way. For self-referencing also, ratio values are to be preferred because a country's stocks of human and financial resources change over time. Such change reduces the meaningfulness of many comparisons based on absolute values. In research applications, it is certainly possible to analyse absolute values because a size dimension can easily be removed from a pattern. Such a methodological approach does not however overcome the problem because it also removes any chance of identifying the relative effort made by a community or a society to provide resources to the education system.

---

[1] The main exceptions to this statement are provided by those methods of indicator formation yielding an absolute value which can be interpreted in norm-, self- and criterion-referenced ways (eg the Gini coefficient and the C index of concentration).

Measurement of stocks and flows

A fourth distinction between types of indicators is that between indicators of educational stocks and indicators of educational flows. The former relates to the number or amount of a certain quantity at a particular time; the latter relates to the rate at which a quantity is changing over time. Data describing rates of change in the value of stock variables are generally reported in percentage terms on an annual average basis – say +1.2 per cent and –3.4 per cent per annum. A positive sign indicates a percentage increase while a negative sign indicates a decrease. Most writers on educational statistics advocate the use of both stock and flow data. Stone, for example, writes: 'stock statistics . . . are of only limited value . . . [and there are] inevitable inaccuracies . . . even in the most straightforward cases, if one has only stock statistics to go by' (1971, 41-2). Other writers recommending the use of both types of statistics include Harbison and Myers (1964), OECD (1967a) and Chesswas (1969). An empirical demonstration of the value of using flow statistics to compare education systems is provided by Johnstone (1977). There, 11 variables measuring educational flows are shown to complement a number of stock variables in defining five distinct patterns of variation among national education systems. A further 12 flow variables combine to form into four additional yet independent patterns which describe other important patterns of variation among national systems.

It is unfortunate that flow statistics are neither reported nor used more often. In some data tabulations made by the UN and Unesco, rates of change in particular variables are indicated but these are infrequent reports and they are generally applied to variables measuring absolute quantities rather than ratio quantities. Gross enrolments and total finance are two typical variables on which flow data are sometimes reported. Rates of change in these variables are of limited usefulness because they are so significantly influenced by other factors (such as birth rates or economic growth rates for the above examples).

Measurement of overall level and distribution

A final classification scheme for education system indicators is based on whether an indicator measures the overall level or the diversity of a characteristic within a country. The majority of indices measure only overall level and this generally represents a kind of average level. Such indicators are referred to here as being 'non-distributive' indicators. Levels of many characteristics vary within different parts of a system, however. Students of varying ages can be enrolled in a particular grade level for example; female participation rates in developing countries often tend to be relatively high in the main provinces but much lower in the more remote provinces. As average values do not reflect such diversity, there is a need for some characteristics to be measured by 'distributive indicators'. These do measure the distribution of a characteristic throughout a system.

Table 1.3 demonstrates the difficulty which can sometimes be met in attempting to make comparisons by using only non-distributive indicators. It reports the percentage of senior secondary pupils who are female in each of the five broad geographical areas of Indonesia. Examination of the values for the level of this characteristic (ie the arithmetic means calculated across the provinces in each area) shows that the East Indonesian provinces are considerably lower in their average percentage while there is little difference between Sumatra and Kalimantan and little difference between Java and Sulawesi. The statistic measuring the dispersion of the percentage values within each area slightly modifies the last part of this interpretation, however. It shows that the Java provinces are fairly homogeneous – the standard deviation value being very small indeed. For the Sulawesi provinces, the dispersion in the percentage values is exceptionally large. Hence some of the Sulawesi provinces have quite low percentage values while others have very high values. To conclude that Java and Sulawesi are similar in respect to this characteristic is therefore incorrect. Statistics measuring average levels should be accompanied by statistics measuring dispersion and both types interpreted in conjunction with each other.

Geographical area	Number of provinces	Level of characteristic (arithmetic mean)	Dispersion (standard deviation)
Java	5	35.28	2.41
Sumatra	8	32.36	3.77
Kalimantan	4	32.13	3.83
Sulawesi	4	34.00	11.04
East Indonesia	5	25.60	4.37

Source: Statistik Persekolaban 1974

Table 1.3: *Statistics describing the percentage of students who are female in Indonesian senior secondary schools – 1974*

The methods for classifying indicators which are discussed above are not intended to provide an exhaustive list. There are many other approaches which have been proposed. In particular situations, Lazarsfeld's (1959) distinction between expressive and predictive indicators could offer additional guidance, as too might Horn's (1975) distinction between subjective and objective indicators. The classification schemes described in detail above, however, are considered to offer a basis for discussing the construction and use of indicators of education systems.

1.4 Additional advantages and disadvantages of developing indicators

In conjunction with the discussion in Section 1.2 on the uses indicators have in the planning and study of education systems, a number of distinct advantages of developing indicators were also claimed. In addition to these, other points need to be mentioned as well as pointing out that adopting a set of indicators can have certain disadvantages also.

The first of the additional advantages is that the development and use of indicators facilitates discussion about education systems. Indicators permit systems to be compared in a norm-, self- or criterion-referenced way across a number of exceptionally complex concepts. As each facet of these concepts is measured by a single variable and these facets (or variables) are combined into an indicator, a composite picture of each concept can be conveyed by the indicator value calculated. On this point, Lazarsfeld and Barton see that 'the special advantage of an index is that it allows us to rank objects in terms of complex concepts' (1951, 181). A corollary of this statement would imply that indicators give an intelligible and comprehensive view of the state of an education system. Presentation of this view is achieved by combining data about specific aspects of systems into an appropriate format. Such a format is preferable to the necessity of having to examine a variety of single variables — the relationships between which are unknown, unidentified or not appreciated.

A second potential advantage of developing education system indicators is that a variety of information, which may not always be strictly comparable, can be synthesized into a form which is sufficiently interpretable for the purpose at hand. This is not to imply that data can be combined to form an indicator without regard to the errors involved in the component variables or without attempting to ascertain the extent of the comparability of component variables. It is instead an admission that variability in the components of an index can sometimes be tolerated more by combining those components with others to form a composite than by examining each individual component in isolation. Unfortunately what has occurred too often in the past is that variables have been left by themselves because of alleged errors of measurement and indicators measuring broad concepts of education systems have thus not been formed. The result of this action has been a growing accumulation of data which is both unorganized and not easily used in presently adopted reporting practices. If data accumulation in this form continues to be the norm, Dahl forecasts that 'we shall be inundated by the coming deluge of information' (1967, 49). Indicators offer a way in which to reduce the information explosion to manageable proportions especially for the purpose of planning, monitoring and researching the development of education systems.

The acceptance of indicators must not be without caution, however.

Disadvantages can exist along with the advantages. One of the greatest of these is the difficulty of interpretation. Sometimes a problem can arise in identifying exactly what concept or characteristic is being measured. Of course this problem can also arise with individual variables but it is compounded when indicators are formed either without a clear definition of the concept being measured or without any regard to the errors of measurement contained within each variable. When indicators are defined without due consideration of the issues underlying their definition, interpretation of their value is rendered impossible. Indicators must be formed in proper ways and the limitations of any value calculated for a particular index recognized. Too often, unfortunately, such care and regard are not forthcoming. Errors such as accepting very small differences between indicator values as demonstrating an educationally significant difference are then committed.

A second disadvantage often attributed to indicators is that they can become a goal in themselves. As indicators for some characteristics are easily defined and measured, development planners might concentrate more on those aspects whose achievement can be easily demonstrated than on other aspects whose definition and measurement are not so straightforward. This alleged disadvantage, often attributed to indicators in other fields of the social sciences, is more a condemnation of the attitudes of the planners than it is of indicators themselves. Certainly, focusing particularly on easily measured features can be a problem hindering the development of indicators. In connection with education system indicators, it has not yet really become a disadvantage. Indeed, as is shown in Section 1.2, the situation has sometimes been exactly the reverse. Goals have been formulated in exceptionally vague terms without any regard at all to the measurement of their attainment. The potential disadvantage must, however, be noted.

1.5 A framework for defining education system indicators

In order to discuss the ways in which education system indicators might be constructed and used, a frame of reference must first be established. Without such a basis, the type of indicators needed in particular situations might not be consistently identified; nor could the conceptually distinct concerns of education system indicators be validly differentiated. The incongruent effects of not working from such a framework have been noted before. Coleman, for example, posits that 'quantitative measurement, to be theoretically fruitful, had better be carried out in direct conjunction with the theory within which it is intended to be used' (1964, 431). The use of a frame has the additional advantage of revealing those areas which at present lack data. One alternative to this approach is to work from the indicators and variables presently available on the assumption that these provide a reasonable coverage of the space referred to as an 'education system'. As such an approach is tenuous in the extreme the former approach is adopted here. This course of action also acknowledges the troubles associated with previous indicator usage for the planning and study of various societal sub-systems. Drewnowski notes the following:

> What has been created so far is an incoherent maze of variables, the definitions of which are muddled, quantification procedures questionable and practical uses, if any, extremely doubtful . . . what is badly needed is the establishment of some ordering principles which would make possible a selection of useful indicators and rejection of the ill-conceived and inapplicable ones. (1972, 78)

Definitions of education systems provide little guidance on the construction of such a framework (see, for example, Unesco [1955], OECD [1967], Chesswas [1967] and Stone [1971]). Indeed, the Unesco *World Survey of Education I* indicates that ' an inclusive definition acceptable in all countries would serve very little purpose . . . [because] no two countries have education systems identically composed and indeed any single system is constantly changing and growing' (1955, 32). Instead of working from such definitions as have been proposed, therefore, a model previously established for empirical research is further refined and developed (Johnstone, 1977).

Diagram 1.1 depicts this model. The main focus is the education system and so the diagram shows this system bridging the gap between a past and a future frame of society. This is admittedly a biassed view of the role an education system plays in the development of a society. A political scientist or an economist would replace this emphasis on the education system with an emphasis he considered of more importance. Nevertheless the bridge adopted here does serve the purpose at hand.

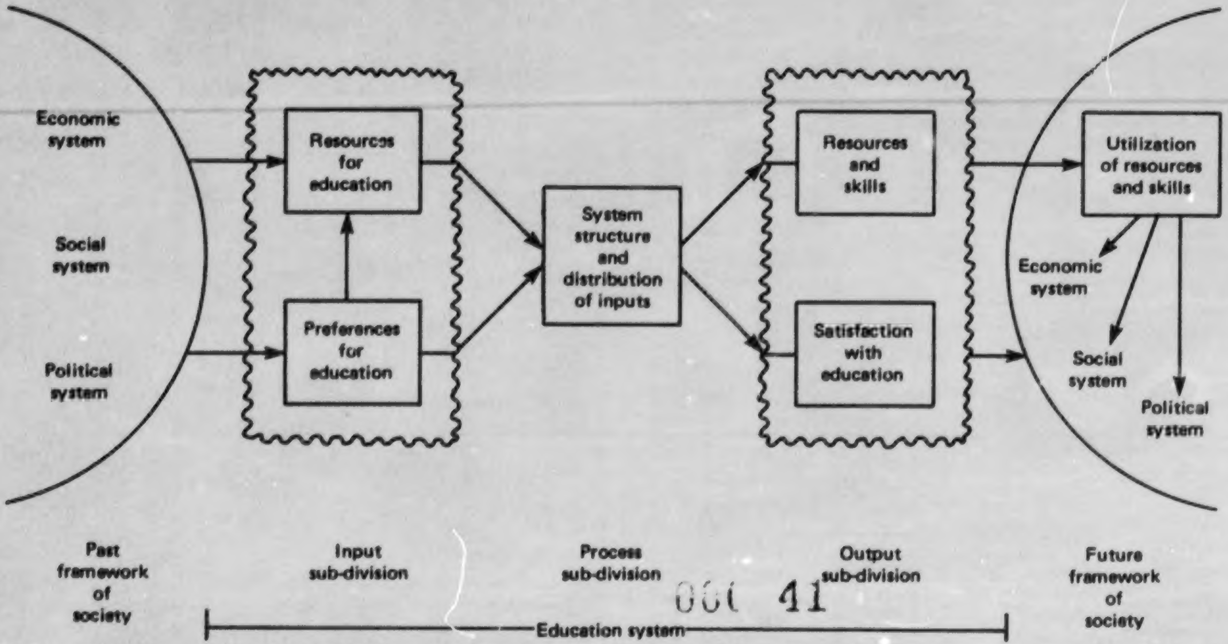
Within the part covered by the education system, three distinct sub-divisions are indicated. These are termed the input, process and output sub-divisions of the education system space. They are considered to be quite distinct yet together they form a basis for a complete definition of the system. The three sub-divisions are thus necessary and sufficient to form the framework required. This model is explicitly based on definitions found within systems analysis theory. Its adoption thus accepts the recommendations of a number of writers[1] in the education systems field who advocate the application of systems analysis principles to the consideration and analysis of education systems.

Although a number of definitions have been suggested for the names given to the three sub-divisions (see, for example, Andrew and Moir [1970] and Banghart [1969]), these are either of indirect usefulness in a discussion focusing on education systems or they contain statements which allow equivocation in the classification of certain indices. Hence

---

[1] Such writers who have advocated this application include Coombs (1968), Love and Kyle (1968), Adams (1970) and Kartomo and Heckelman (1971).

40


Economic system
Social system
Political system
Resources for education
Preferences for education
System structure and distribution of inputs
Resources and skills
Satisfaction with education
Utilization of resources and skills
Economic system
Social system
Political system
Past framework of society
Input sub-division
Process sub-division
Output sub-division
Future framework of society
Education system

000 41

Diagram 1.1: *A framework for the development of indicators of education systems* the definitions used previously in the empirical research study cited above are adopted for use here:

- ☐ Educational input indicators are either indicators relating the amount of a particular quantity taken by an education system to the total amount available for distribution or indicators describing the aspirations held by a society for education systems.
- ☐ Educational process indicators are those indicators describing the structure of the system which processes the inputs to become the outputs or those indicators describing the distribution of the inputs throughout the education system.
- ☐ Educational output indicators are either indicators relating the amount of a particular quantity leaving an education system to the amount with some similar characteristic which is available to leave or indicators describing the perception by a society of the results of education systems functioning.

In this defined framework, educational input indicators are totally exogenous in nature as too are those process indicators related to structure. Both of these types of indicator are thus determined by decisions taken outside of the frame. This is not ignoring the fact that certain process or output indicators may not affect policy decisions and that these may in turn affect the value of those indicators. It assumes instead that the influence is not direct.

This perception is quite different to that held for the output indicators. These are considered to be totally endogenous in nature. They cannot, for example, be directly affected by policy decisions except to a fairly minor degree and even this would be in the most unusual circumstances. The values taken by output indicators are thus determined by the values for input and process indicators.

The division of the education system space as proposed can also be viewed in a purely mathematical way. It is possible to achieve the same concepts as are intended in the above definition by writing the simple equation:

$$O = f(I)$$

where the symbol O represents the set of output indicators
I represents the set of input indicators
f represents a function of the process indicators used to convert the input quantities to outputs

Some may criticize the concept of dividing the education system space into these three sub-divisions by claiming that a single indicator (or variable) can be classified as being an input in one sense, a process in another sense and an output in yet another sense. For example, the ratio of enrolments at the second-level of an education system to the enrolments at the first-level might be regarded as:

- ☐ an input variable because it indicates the relative quantities of pupils who can be expected to enter either the work-force or third-level education in the next few years
- ☐ a process variable because it indicates the relative distribution of enrolments across levels
- ☐ an output variable because it indicates the proportion of students who have recently benefited from the first-level education and who have remained within the system compared to those who are presently benefiting.

As conceived within the proposed framework, the classifications 'input variable' and 'output variable' are incorrect. Input variables, as defined here, must relate educational quantities back to another societal subsystem (for enrolments, this could be the number in the population). Output variables must relate quantities leaving the education system to similar quantities still within the system (for example, graduates to enrolments at a particular level). One major implication of the proposed framework, therefore, is that each indicator can and should be classified into only one sub-division of the education system space.

Input sub-division

Within the input sub-division, there are two separate considerations for indicator definition and selection. One relates to resources while the other relates to preferences. The first type of input — Resources for education — needs little explanation. Implied here is the proportion of the relevant physical resources, such as manpower and finance, which is devoted to providing the system of education in a country. Examples of typical variables here include enrolment ratios and the percentage of the total national budget devoted to education. These variables are similar to the sorts of variables Poignant lists when he calls for 'discovering a set of criteria that would embrace the means devoted to education assessed with reference to national resources as a whole' (1972, 28).

The second type of input — Preferences for education — is conceptually very different. Indicators classified here do not reflect the physical quantities themselves in any way. They reflect instead the preparedness of a society to devote resources to education as well as the perceptions held by the society to the provision of an education system and the aspirations held for such provision. Indicators of this type therefore reveal the extent to which education is preferred to other nationally provided facilities such as health services, social benefits, defense services, etc. For example, a country could not have such a preference for education that its health services and facilities have to be terminated. Instead there comes a point when a country must decide, in some way, that additional expenditures in one field can only be justified after expenditures in other fields have been made. Of course that country may also consider that expenditures in no field should be allowed to fall below a certain level. The preferences for education must therefore be determined from the relative priorities across the different benefits stemming from the margins of investment in each facility.

Preferences for education are mainly expressed by two sectors – the political sector and the private sector. These represent respectively those who provide and those who directly consume education. Generally there is a close agreement between the preference expressed by both groups although there does not necessarily have to be a congruence. In most democratic countries at least, the preferences expressed by the two sectors can be formed relatively independently. In countries having different political systems, the process of forming preferences can be more complicated. It might, for example, be a very different process or it might not be as evenly balanced.

The level and nature of preferences for education within a country must affect the level of resources input into a system – hence the arrow in Diagram 1.1 connecting these two boxes. If both governments and individuals have strong preferences towards, for example, second-level education, enrolment ratios for this level should be high. If, however, third-level education is provided by a government but individuals see little benefit in it for them, third-level enrolment ratios would be low.

Analogies can be drawn between this type of education systems indicator and indicators in other areas. Culyer *et al* (1972) define an indicator in their study measuring 'need for health'. Drewnowski and Subramanian (1970) discuss the 'revealed preferences' of planners in development plans while Drewnowski (1970) argues for the use of preferences to be incorporated into a 'level of living' index.

Process sub-division

The process sub-division needs little further comment. As the definition states, indicators here reveal the distribution of quantities throughout an education system as well as the structure of the system. The one-way left-to-right arrows into the Process box in Diagram 1.1 represent the requirement that process indicators are affected by, but cannot themselves directly affect, the values of input indicators.

Output sub-division

There are two distinct considerations within the output sub-division of the education system space. In the first of these – Resources and skills from the education system – indicators related to the quantity and level of skills produced by an education system are defined. These describe, for example, the relative numbers of students who graduate from a particular level or the cognitive ability of students in particular subject areas.

As the second type of output, indicators related to the level of satisfaction with the education system are defined. In part, such indicators reveal how effective a system has been in allowing individuals to develop their own skills and abilities. The indicators also reveal the degree of satisfaction of such groups as employers, with the skills possessed by students who leave the education system at different points.

Education system – societal congruence

The final component depicted in Diagram 1.1 measures the congruence between the outputs of the resources and skills from the education system and the needs of the society into which the outputs flow. It is through these measures that the effects of the education system on a society are to be noted. Indicators here consider such factors as unemployment rates, appropriateness of skills, effect of education on income, improvements in health levels (as a direct result of health campaigns in schools) and literacy rates. The external efficiency of the system can thus be assessed by indicators developed at this stage. If these indices reveal that an education system is not contributing to the society which is providing that system in the way it was planned it would contribute, policy decisions related to varying inputs to and/or processes of that education system would need to be taken.

Mushkin and Billings (1975) posit that the last two components of the education system space can be elaborated further. By dividing outcomes into general types such as personal and social outcomes, they propose another way of expressing what is maintained here. They also add another dimension to their model by claiming that educational effects accrue over a long time period. Some benefits are immediate, others are secondary or accrue shortly after formal education is complete, while yet others are inter-generational and will, in turn, form the nature, values and perceptions of society in the future. The time differential they propose is quite proper. The first two parts of the time scale are implied in the framework described here but the third cannot be so incorporated. Over a long time period such as they envisage, too many other influences can affect attitudes towards education. Hence output indicators are still considered to be independent of input indicators – an assumption which could not be supported by accepting the Mushkin and Billings proposal.

The model advanced in this section is applicable both to education systems viewed on a national scale as well as to education systems viewed on a regional level. The considerations involved are identical – the only difference is in the definition of some of the indicators and/or variables used within systems at either level. For example, the variable 'percentage of the gross national product spent on education' is not applicable to a provincial or a state education system although an analogous variable might be definable. Adopting a frame capable of covering a variety of levels has the advantages of maintaining comparability between indicator sets where this is possible as well as facilitating the development of indicators at one level of complexity through allowing the adoption of concepts underlying indicators developed at the other level. Both of these advantages help to provide a common basis on which to examine education systems. Unfortunately such a basis is too often lacking especially in studies designed to evaluate education systems provision. National concerns are seen to be completely different from the concerns experienced at state or district level and no overlapping of concepts is seriously considered to exist.

Implications of the theoretical framework

The theoretical framework as suggested here has four immediate implications for the definition of education system indicators and the estimation of values for these indicators. One implication is that no single index could or should be estimated to represent education systems. As is indicated previously, indicators must be composite in nature. The framework therefore implies that aggregation of either variables or indicators must not continue so far as to form a single broad education system indicator. Suggestions advocating such an index are, however, frequently made. Perhaps drawn by such comfortable 'stars' as gross national product and IQ, many writers and researchers advocate the formation of one indicator for education, another for health, another for urbanization and so on. Other writers even argue a total composite should be formed across broad levels of society to form an index of the level of living or welfare.[1] Fanchette describes this desire as being 'a quest for some sort of "Gross Happiness Product"' (1974, 9).

When gross aggregation occurs, many difficulties arise. The most outstanding problem is that of interpretation. To demonstrate this, an (improper) assumption is made that it is possible to form a single indicator for each of the five educational areas depicted in Diagram 1.1. The values for two countries – call them country A and country B – on these five indicators are then estimated. Diagram 1.2 depicts their profile on some arbitrary, common scale. The countries are shown to be very different from each other in all of the various aspects of their educational provision. An 'index of education' is now developed by calculating an average value across the five indicators.[2] This average level is also indicated in Diagram 1.2. Although the 'composite index' shows the two countries have similar levels of education, the profiles show such a conclusion is invalid. Hence aggregation across distinct areas within the field of education alone confounds the interpretation

---

1. For discussion of these broad indices see, for example, Ferriss (1969), Drewnowski (1970), Kamrany and Christakis (1970), Stake (1970), McGranahan *et al* (1972), Adelman and Morris (1972) and Yeh (1974).
2. The index could also be formed using some complex weighting procedure but the point of the discussion remains valid.

of a composite index. When aggregation occurs across areas which are very different in kind to form an index of the level of living for a society, interpretation becomes impossible. In contrast, under-aggregation of variables loses much of the advantage of forming composite indicators. A semi-aggregative indicator would not necessarily ensure that reliability or internal validity are as high as they really should be. This issue is taken up further in Sections 2.4 and 2.5.

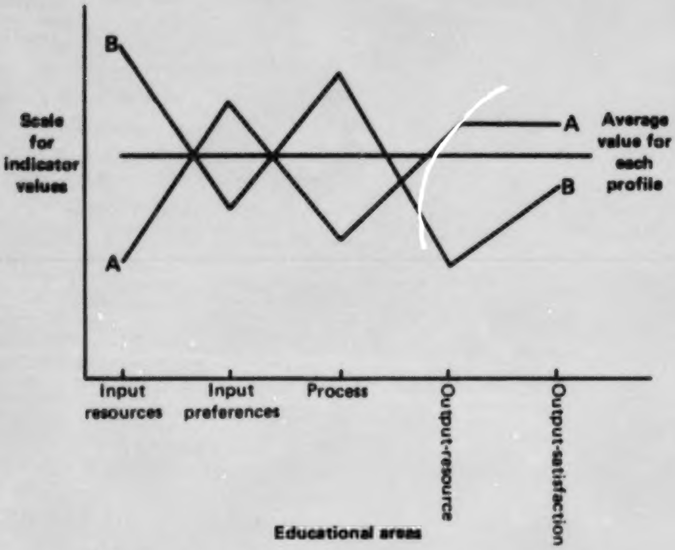


Diagram 1.2: *Hypothetical profiles for two countries across the five educational areas*

A second implication stemming from the adoption of the theoretical framework is that indicators of education systems should not measure only learning outcomes. Certainly development of achievement levels of individuals is the ultimate purpose (in one form or another) for education system provision. In the planning, evaluating and researching of education systems, however, the analysis of achievement levels must be only one aspect of a wide spectrum of considerations. Indicators of inputs and processes are of equal importance to indicators of achievement level. So too are indicators of other kinds of output. All of these kinds of indicators comment on particular concerns which are conceptually, and often practically, distinct. Hence, focusing solely on academic levels to plan or evaluate the performance of an education system takes far too narrow a view.

A third implication of the framework is that some of the data required to construct indicators cannot be collected as part of the normal educational administrative procedures. Many data will instead have to be collected via survey methods. Although such procedures are most unusual in the educational context, they are not unusual in the collection of economic data. In that field, for example, a number of statistical series can only be determined by interviewing samples of particular groups within the community. (See, for example, many of the series reported in *Social Indicators* and *Social Trends*.) The adoption of similar procedures for acquiring education system data is thus advocated here.

A fourth implication relates to the focus of possible theory development. The framework establishes the education system as being a substantive field for study and one that is in many ways distinct from the society in which the system is located. Two sorts of theories might therefore be developed. One of these is a theory of education system development. It would describe and predict the way in which a system evolved. Such a theory, if it existed, would help administrators and planners identify particular types of decision which might facilitate the making of desired system changes in more efficient and effective ways.

The other sort of theory is a theory of education system effects. This would describe and predict the impact of the provision and development of an education system on the other sub-systems of society – such as the economic, political and social sub-systems. The studies listed in Table 1.1, for example, all hypothesize results such a theory might encompass. The two sorts of theories identified here, if they do evolve, would be developed fairly separately if the framework proposed is valid. Empirical testing therefore needs to be attempted through the development of indicators to determine if this assertion is plausible.

1.6 Considerations to make when constructing indicators

There is no already developed and accepted statistical procedure for developing indicators in any of the fields which contribute 'social indicators'. Methods for developing indicators are really still very much in their infancy. This is in marked contrast to the situation in economics where indicators for particular phenomenon have been developed and their characteristics investigated for some considerable time.

Most of the approaches to indicator formation address four major decisions:

- ☐ which method of definition is to be adopted?
- ☐ which variables are to be used?
- ☐ what method of combination is appropriate?
- ☐ what weights are to be used?

Within the first of these decision areas, it is possible to identify three broad methods of definition. These are:

- ☐ pragmatic reduction
- ☐ theoretical definition
- ☐ empirical definition

Very often the adoption of one of these methods of defining indicators dictates the appropriate course of action to be followed for the other three decisions. Hence the general issues concerned with the decisions to be taken in indicator formation are now discussed within the context of these three broad methods. All four decision areas are elaborated at length in subsequent chapters.

The pragmatic definition of an indicator

Pragmatic definition of an indicator occurs when a selection is made from a number of available variables or when a number of available variables are combined because of certain assumed (not demonstrated) relationships among them. In the former mode, the type of indicator formed is really the representative index such as was discussed earlier. In the latter mode, the complexity of a defined space is reduced perhaps because it is desired to decompose 'certain groups or combinations . . . to one class in view of [a particular] research purpose' (Lazarsfeld and Barton, 1951, 174). The principal feature of such combinations as characterize this approach to indicator formation is that they are based purely on the judgement of an individual. They thus depend upon the attitudes (and biasses) of that individual in selecting one variable or attribute rather than another. Because of such dependence, this method is considered to be a very weak method of index formation. Actual and demonstrated relationships rather than personally assumed relationships must be used. This is not to deny that the other methods for indicator definition do involve judgements of some kind. What is different here however is that the whole method is judgemental and it needs no empirical support whatsoever. At this time, our understanding of education systems is not sufficiently well advanced on these grounds to make such all-encompassing judgements. Pragmatic definition as a method of indicator formation is therefore not discussed further.

The theoretical definition of an indicator

The method of theoretically defining indicators involves the arithmetic combination of a number of variables which have been defined *a priori* as being inter-related. Each of these variables then has a numerical weight assigned to it, the values for which do not depend on the actual data being used. The choice of the particular variables to be combined may be based either on theoretical grounds or on an attempt to define a concept integral to a development plan. For example, an indicator I might be defined *a priori* to be composed of four variables – say V_1,

V_2, V_3 and V_4. In addition these four variables are to be combined through the basic mathematical operation of addition. To form I, therefore, the four variables must be weighted in some way so that

$$I = a.V_1 + b.V_2 + c.V_3 + d.V_4 \qquad \textit{(Equation 1.1)}$$

If all of the variables have the same perceived importance (and all four distributions are equated so as to have the same variances – see Section 3.2) then

$$a = b = c = d = 1$$

Adding together the values for the four defined variables produces the final value for indicator I.

If, however, V_1 is perceived as being twice as important as V_2 which in turn is three times more important than V_3 and V_4 (which are perceived to be equal) then

$$a = 6;\ b = 3;\ c = d = 1$$

and the indicator is calculated as

$$I = 6V_1 + 3V_2 + V_3 + V_4 \qquad \textit{(Equation 1.2)}$$

The selection of variables and their weights requires extreme care and proper consideration. If these are not forthcoming, not only will the numerical value of the indicator change but also the implication of the concept being measured and, quite possibly, the purpose for which the indicator might be used. Lazarsfeld and Barton cite Zeisel's discussion of baseball indices to demonstrate the problem which exists with using different weights. They note that 'batting average measures the frequency of any sort of safe hit, while an index weighting each hit by the number of bases indicates the power of the hitter as well' (1951, 183). Here the different weights change the implications of the underlying concept so that two very different indicators are formed. The two indicators are both interpretable in proper ways but they have different purposes in the context of analysing baseball games.

Weights for the component variables can be selected in a number of ways. Judgement of the relative importance of each variable by individuals either directly or indirectly involved in constructing the index is by far the most common approach. This method as well as certain others is discussed in Section 3.1. Suggestions from policy documents can also be used. Drewnowski (1970) and Drewnowski and Subramanian (1970), in arguing for the construction of a welfare index, posit that weights for such an index could be derived by considering different welfare needs. This method partially overcomes certain arguments levelled against the arbitrary nature of some of the methods for deriving weights. Much work remains to be done in this area however.

There is also a third aspect involved in the theoretical definition of an indicator. This is the selection of the actual method of combining the defined and weighted variables to form the indicator. There are many variations which can be used to specify an indicator in mathematical terms. By far the most common method involves simple addition. Both of the above equations, for example, specify an addition definition for the indicator. The other common basis for the indicator specification is the multiplicative model. An example of such a model would be

$$I = a.V_1 \frac{b.V_2}{c.V_3} \qquad \textit{(Equation 1.3)}$$

Such a model may occasionally include variables raised to powers. The distributions of indicator values derived from such a model are very different. Bowles (1969, 63-72) shows such differences most convincingly. He calculates values for four different indicators of labour requirements for a group of 12 countries. The correlations between pairs of indicators across his sample range from 0.54 to 0.99. This result shows that quite different conclusions could be reached depending on which indicator specification is selected. It is very important therefore to identify the sort of model required for a particular indicator.

The essential difference between an additive and a multiplicative specification is that a simple additive model assumes a perfect substitutability in values among the variables included in the definition. For example, for equation 1.2 in this section, values of V_3 could substitute directly for values of V_4. In contrast, a multiplicative model implies that the substitutability amongst components is either zero or at least calculable and considerably less than 1.0. Further discussion on this point is deferred until Section 3.1. Suffice it to state summarily here that, in the theoretical definition of an indicator, the three problems to be addressed and solved are:

- ☐ the selection of component variables
- ☐ the selection of weights
- ☐ the selection of the method by which the variables can be combined into the composite indicator.

The empirical definition of an indicator

The third method of forming indicators is that of empirical definition. This method can be viewed as an analogy of the theoretical definition procedure just discussed. Its distinguishing feature is that, instead of selecting weights in an *a priori* manner, weights are here determined by analysing a set of data and identifying strengths of the functional relationships existing among the variables. Methods such as factor analysis, cluster analysis and Guttman scaling are particularly appropriate here. They are discussed in detail in Chapter 4. When factor analysis is used to analyse data, for example, loadings for each variable are estimated on each factor. These loadings thus become the weights to be used for each variable in forming a score for each factor extracted. The factors can be regarded as the indicators.

In forming indicators by empirical means, a major problem exists in that the data and not the underlying constructs could determine the character of the indicators. Such a bias can be introduced either by the selection of variables used or by the cases (eg the countries or regions) across which the analysis is performed.

Variable selection is notoriously bad in studies seeking either to develop indicators or, more generally, to identify dimensions. Very often no justification for the selection of a particular variable set is given and an *ad hoc* group of variables happens to be analysed.[1] The possible results of this approach are well demonstrated by Cant (1975). In attempting to identify variations 'in population structure, types of livelihood and levels of living in rural New Zealand' (p71), Cant selects 28 variables but gives no justification for his method of selecting these variables in particular rather than others. He posits that, for his problem, 'differences in the age and sex structure of rural New Zealand communities' (p72) are important and hence he introduces 'an unusually large proportion of demographic variables' (p72). It is not surprising then that the subsequent factor analysis results in four of his five factors manifesting a predominantly demographic structure. Further complicating his analysis are the almost random associations which occur as a result of his selection procedure. For example, his indicator (factor) of 'Farming type' describes a community having many children in the 10–19 year age group, relatively fewer males in the 25–44 years age group (compared to other communities) but having small farms and many dairy cows and cattle. The implication of a positive relationship among young children, adolescents and cows is unclear and probably results more from Cant's *ad hoc* selection of variables than from some construct underlying the nature of rural New Zealand communities. Almost every study analysing variables chosen neither according to definite criteria nor within a pre-determined framework yields indicators of limited or dubious value.

The other aspect of this methodological problem encountered when forming composite indicators by empirical definition is that the cases themselves can determine the values of the weights. It must be expected, for example, that if data are collected for all countries across a set of variables, separate analysis of the following country sets and sub-sets would certainly yield different weights – and probably even different indicator sets:

☐ all countries together
☐ only African and Asian countries
☐ only member countries of the OECD.

---

[1] To demonstrate this point, see studies such as those by Schnore (1961), Cutright (1963), Harbison and Myers (1964), Gregg and Banks (1965) and OECD (1971).

If educational variables were to be analysed, it could be hypothesized that variables related to first-level education would be more important in set 2 than in set 3 but that third-level education variables would be more important in set 3 than in set 2. These differences would yield different indicators. (For a further example of this point, see the differences in the correlation matrices calculated across various country sets and reported by Harbison *et al*, 1970, 118-21.)

Despite these two problems of variable and case selection, the empirical determination of weights does appear to offer great promise for indicator development. If variables and cases are chosen methodically and purposefully, at least a basis is formed for further development. But the essence of theory development – replication – must also be undertaken, particularly to determine if indicators identified at one point in time are still valid at another point. McGranahan *et al* warn that cross-national patterns are not necessarily stable cross-temporally (1972, 5). Adelman and Morris (1967, 144) and Peaslee (1969) also allude to instability as a possible problem of indicator development. If fairly consistent associations and weights are identified through a number of replication studies, however, indicators having both reliability and validity can be accepted. Indeed, in such studies, the technique of 'successive definition' as developed by Adelman and Morris (1967 and 1972) can be used with advantage especially if the empirical and the theoretical methods of defining indicators are employed at different stages.

A third issue raised by the empirical definition of indicators is the determination of which variables can be aggregated to form an indicator. This issue has been discussed in part in conjunction with both of the previous issues. Certainly for the estimation of levels of education in a country, indicators must not be defined in conjunction with other sub-systems of society. Education system indicators must be developed quite separately from health or urbanization indicators to name just two other areas. Indeed it is recommended here that variables in the various sections of the sub-divisions of the education system space – as depicted in Diagram 1.1 – must not be combined. So, for example, variables related to the two kinds of inputs should not be combined. Similarly, but more definitely, input and process variables must not be aggregated into the one index. Within an area, no recommendation can be made in general terms. If all variables reflect the same type of educational feature then they can be combined. Within the process sector, many different dimensions can be identified and hence a number of composite indicators should be formed there.

Chapter 2

Statistical Issues Deriving from Education Systems Data

One of the major obstacles hindering the development of the study of social systems has been the problem of quantification. This problem has been manifest in the almost wholesale acceptance of the scales and measurement principles from the theory of psychometrics. A large part of this acceptance has been most proper for there are problems in the study of social systems which are in essence the same as those found in psychometrics. Some of the acceptance cannot however be justified. Social systems have some properties which are either different in nature or different in kind from the properties of individuals. To adopt the principles of psychometrics unquestioningly can only lead to difficulties or errors both in analysing the data derived from social systems and in interpreting the ensuing results. Before collecting data from social systems, therefore, variables must be appropriately conceptualized and defined. The whole measurement operation indeed needs very careful planning and execution.

Two of the most basic and important issues to be considered when conceptualizing and defining the variables on which education systems data will be gathered are:

- ☐ to what units will the data refer?
- ☐ what is the theoretical nature of the scale which underlines each variable defined?

Once a position has been taken on each of these issues, some of the more commonly perceived data problems of validity and reliability also need attention. This chapter outlines some of the main issues involved with the process of conceptualizing variables which can be appropriately used for the measurement of phenomena in social systems. It also discusses the definition and collection of data for educational planning purposes on those variables from education systems in particular.

Before commencing this discussion, a caveat must first be issued. Data can be collected for two purposes. One of these is for planning; the other is for research. With regard to many considerations, the nature of the data and the demands made on them by these purposes are identical. This does not always have to be the case. When there is a difference, the discussion in this chapter concentrates on the considerations related to data collection for planning purposes.

2.1 The nature of education systems data

Individual or collective units of measurement

The first issue which arises when considering the nature of variables for an education system and how these variables might be measured is the definition of an entity or unit to form a basis for the observations to be taken. A selection must be made for each variable between one of two types – individual units and collective units. The former describes a single entity – generally a person – and it is the unit used most in psychological measurement. Age, sex and number of years spent in school are typical examples of variables measured on an individual unit. The latter type relates to a particular group or set of individual units taken together. In some instances, this collective could be a class group, while in others it could be a particular school or even an education system aggregated to either a provincial or a national level. Variables such as class size, number of years of education provided at the first-level and the percentage of the teaching force who are female are typical variables defined on collectives rather than individuals. In general, the entity being measured must be carefully defined so as to allow an appropriate inference from data analyses.

Part of the problem associated with the definition of an appropriate entity to be measured is manifest in what has been called the 'ecological fallacy'. Robinson (1950) was the first to demonstrate this. He examined two variables – colour and illiteracy – and inter-related them across entities at different levels of aggregation. Measured across large census sub-divisions of the US, the correlation between these two variables is 0.95. Two conclusions which may be drawn from this value are:

- ☐ an extremely strong relationship exists between the percentage of the population who are coloured and the percentage of the population who are illiterate in the US
- ☐ individuals in the US who are coloured also tend to be illiterate.

Recalculating the correlations on the basis of states rather than census sub-divisions, Robinson obtained a value of 0.77. This therefore weakens the first of the above conclusions. The correlation was then recalculated again – this time based on individual units. Here the correlation is between a variable indicating whether a particular individual is coloured or not against whether that person is illiterate or not. Robinson reports a value of 0.20. Such a low value completely revises the second of the above conclusions. Indeed, both of these conclusions demonstrate the ecological fallacy of generalizing from the behaviour measured at one level of aggregation either to another level of aggregation or to the behaviour of individuals.

In a different sort of aggregation, the ecological fallacy can also be demonstrated by the following series of propositions:

☐ country X is in the Asian region
☐ Asian countries have characteristic Y
☐ country X has characteristic Y.

Country X may or may not have characteristic Y. To determine if it does, country X should be examined. In this case, a single country should be used as the collective to be measured rather than the previously used region.

Both of the above examples demonstrate that generalization should not be made from measurement conducted at one level of aggregation to measurement conducted at another level. The statistical reasoning for this is that units related to larger entities tend to be more homogeneous than units related to smaller entities. Hence variation amongst individuals can be expected to be greater than that produced by aggregating individuals located in each of a number of provinces. To adapt values for correlations based on aggregate units instead of on individuals is thus 'to use a statistic which actually measures the "between" relationships to serve as an indicator of the "within" relationships' (Fennessey, 1968, 20). The nature of the difference between these two types of relationship is difficult – if not impossible – to describe in specific terms because the two components are determined independently in any given set of data. Suffice it to note that the differences can bias any extrapolated statistical association. Scheuch (1966, 149) offers a parallel explanation of the phenomenon in terms of the marginal cell frequencies of contingency tables.

The importance of defining appropriate units on which measurement can appropriately occur is also manifest in another statistical problem. This arises when characteristics related to one level are related to characteristics at another level. For example, the relationship between achievement test scores and class size might be investigated. The relationship could be estimated by correlating, across a large number of classes, mean achievement scores in a class with the number of pupils in that class. If a sufficiently large number of classes is not available, however, individual achievement scores might be correlated with the size of the class in which an individual is placed. As a number of individuals would come from the same class, the problem known as multicollinearity occurs (see, for example, Gordon, 1969). The effect of this is to reduce the variability in one variable quite substantially and hence lower the calculated correlation to close to zero. Hence a 'true' inter-relationship is hard to detect when units related to different levels of aggregation are used in the same calculation.

A major implication of being able to define variables measured at different levels of aggregation is that planning and theory development can and should be conducted at different levels. As propositions at one level of aggregation do not necessarily parallel those at other levels, each level needs to be investigated and distinctions between levels taken into account. No implication can be made that one level is more important than another. Ferriss makes a comment, for example, that 'to develop educational statistics into the needed tools for planning and evaluation . . . knowledge first of the inter-relationships among variables in the most elemental unit is needed, interaction in the classroom and the school' (1969, 5). Other writers draw similar distinctions between micro- and macro-theories and imply that micro-theories are the ultimate goal of development for educational research effort. The validity of such assertions as these must be seriously questioned in light of such phenomena as the ecological fallacy. Certainly knowledge about inter-relationships at the individual level or the classroom level of aggregation is needed but this must accompany knowledge about relationships existing at the provincial as well as the national education system level.

A second implication of having different levels of aggregation is that different sets of indices need to be defined for different levels of education systems. Indicators of national education systems, for example, could contain a core which parallels those indicators defined for provincial systems or systems in more localized planning regions. Each of these sets, however, would also contain other indicators not having counterparts in the sets related to the other levels. Research aimed at developing indicators therefore needs to be conducted across a variety of educational planning levels.

Aggregate or global data

Having established that data for education systems measurement should be based on collective units of measurement, a further distinction within units classifiable as being collectives is necessary. One kind of collective data can be referred to as aggregate data; another kind is global data.

Aggregate data measure the features of a system which are derived from the features of those individuals or units making up the system. They summate in some way the individual characteristics and therefore comment on the occurrence or the distribution of those characteristics throughout the collective unit. Typical variables of this type include the percentage of first-level students who are female and the average years of schooling delivered to a population.

In contrast, global data do not depend on the characteristics or the attributes of the individuals who make up the unit of analysis. They instead comment on the features of the unit taken as a totality. Measures of the global type include the number of grades of education provided at the first-level, Farrell's (1969) index of structural differentiation and a variable measuring the degree of central control over matters of educational administration.

The major implication of using aggregate data to describe a social system is that it opens up the opportunity of using a number of distinctly different sorts of measurements. Lazarsfeld and Barton (1951) discuss three of these – a rate, an average and a parameter measuring the distribution of the measurements. It is not necessary to use all three of these kinds of statistic concurrently for any one variable or indicator.

A rate is a measurement of the proportion of members in a collective unit possessing a certain property. Such variables as the percentage of the work-force who are teachers, the percentage of the gross national product spent on education and the percentage of all buildings used for more than one shift per day are examples of rates. Any value calculated applies to the collective defined as forming the unit of measurement but it is estimated by examining the attributes of all of the individual members within that unit. Generally a value for a rate is estimated by regarding a variable as having a number of discrete categories (ie having an underlying nominal scale). For example, for the first of the above variables, the measurement would involve ascertaining 'teacher' or 'non-teacher' for each member of the work-force. For some variables, this is the only sort of measurement possible.

When a characteristic can be measured in a more sophisticated way so that an estimate can be made of the differences in degree to which that characteristic is possessed and not just the kind of characteristic, then both a measure of central tendency of the distribution (ie an average) and a measure of dispersion can be calculated. Very often only the former is reported and this is done by either a mean or a median value. Mean test scores or median years of schooling possessed by a population are two commonly reported education system variables of this 'non-distributive' type. Another measure of central tendency – the mode – is very infrequently used to report educational statistics. It is susceptible to extreme fluctuations when only small changes in a distribution occur. Such fluctuations can severely distort cross-temporal or cross-regional comparisons.

Non-distributive indicators, however, reveal only one aspect of the distribution of members within the collective. Equally important, and mathematically independent of the average value, are measures of dispersion. These measures can be referred to as 'distributive indicators'. The extent to which an average value can be used to describe all members in the collective can now be estimated. The larger a measure of dispersion, the less the average value is truly representative. The statistic 'standard deviation' is a common measure of dispersion and it is generally appropriate to use for most educational variables of the aggregate type. Other measures are also available and these are outlined in Chapter 3.

Unfortunately, few tabulations report dispersion variables or include distributive indicators. Certainly this is sometimes because of the nature of the measurement scale being used (ie it is an ordinal scale – see the discussion in the next section) and a measure of dispersion would be a misleading figure at best in these instances. In other situations, the possibility is simply ignored.

Global data are of an entirely different type. By examining the collective unit as an entity, the same sorts of measures as are discussed above cannot be applied. Instead observations must be made concerning the sorts of characteristics the collective has. As no data need to be obtained from members of the collective, estimates can be made of global characteristics either from policy statements or by expert observation and review.

One aspect hindering the use of global data is that the variables measured tend to be regarded as being more 'qualitative' in nature and hence less amenable to statistical reporting and analytical manipulation. They are thus generally ignored. Instead of reporting data of this very basic descriptive nature, tabulations more often report total quantities (or aggregate data). Sometimes this is very proper because the global characteristics can be inferred when and if required. Indeed when this facility is available, the aggregate data are often preferable because they provide more insight into the actual situation prevailing. For example, reporting the percentage of students in each of the state, religious and non-denominational private sectors of an education system provides more information than indicating whether or not the state-provided education facilities form the main system or not. However, data concerning issues such as how a system is planned and administered and the location of responsibility for curriculum development and assessment reveal equally important issues of what an education system is like. Their less apparent sophistication and complexity generally ensures that global variables like these have no place in statistical tabulations.

A second yet allied observation about global characteristics is that they tend to be the features of an education system most amenable to manipulation for change by policy decisions. Any of the above characteristics given as examples can be altered at any time with what could subsequently be major changes to aggregate variables within the input, process and output sectors of the education system. Ignoring these global characteristics in variable definition or in index formation therefore can severely weaken attempts at explaining inter-system variations or intra-system developmental changes.

No imputation should be made from the foregoing discussion that one or the other of the two types of collective data unit provide a better basis for defining education system variables which might subsequently be combined to form education system indicators. Both kinds measure attributes of the collective and they generally measure different attributes. Hence both aggregate and global data are required.

The only implication which should be drawn is that the essential differences between the two kinds must be recognized and accommodated in any complete indicator set.

2.2 The nature of the scales used for measurement

The issue of scale is of fundamental importance to the whole measurement operation. How a system is to be examined is determined by the scale, as too are the procedures which can be validly performed on the collected data so as to form indicators from it. Scale also imposes constraints on the type and scope of the inferences which can be made by analysing the data.

All measurement scales have two properties:

- ☐ they allow every object being considered to be ordered in such a way that some degree of comparison is possible
- ☐ they allow a system whereby numbers can be assigned to categories on the scale so that comparisons might be facilitated and certain mathematical operations might be performed.

Without these two properties, no method allowing examination of a group of objects could be developed. Any analysis would have to be focused on a single unit – for example, on an individual, a school or a state or national education system. In addition, this analysis would have to be performed at a very low level of description for it would be denied any explanatory power.

The extent to which measurement scales possess these two properties varies quite considerably. In the general measurement literature, at least four main types of scale are discussed. These are called nominal, ordinal, interval and ratio scales.

Nominal scales

Nominal scales are the least sophisticated. They merely facilitate comparisons on the basis of kind but not on the basis of degree. They compare entities as to the type of attribute possessed rather than on the degree to which an attribute is possessed. Indeed, because of this limitation, some writers claim that nominal scales cannot be properly classified as measurement scales. A country may be classified in a particular world region on a nominal scale; its education system might be classified as providing elitist or general education or whether the majority of its funds come directly from local or national governments; its students and teachers can be classified as being either male or female. In all of these examples, two or more categories are available to allow comparisons among the entities on which the measurement is being made. As an entity cannot be assigned to more than one category (provided the scale is properly constructed), conclusions regarding either similarities or differences can be derived.

Normally numerals are not assigned to the categories on a nominal scale because few mathematical operations are formally allowable on such scales. The operations of addition or multiplication normally do not apply. Hence, except for computerizing the data and certain special statistical operations, no formal number system is necessary for a nominal scale.

Ordinal scales

Ordinal scales allow slightly more refined measurements than do nominal scales because they do allow comparisons among entities in terms of the degree to which a particular attribute or trait is possessed. Comparisons amongst entities are thus effected in a very simple way. The amount of the quantity being measured is not reflected by the ordinal scale – simply each rank within the particular group being examined in a measurement operation. Given additional group members, ranks could change. This would happen, for example, if all African countries were ranked on a particular characteristic of their education systems and then all Asian countries were subsequently added to the group. Comments concerning how different one entity is from another or how much more of a characteristic is possessed by one system over another cannot be made when an ordinal scale forms the basis of a measurement operation.

Numerals are assigned to the entities in a straightforward way in that 1 is generally assigned to the entity which has been ranked first, 2 to the next and so on. Numerals are averaged when more than one entity has been equally ranked. The assignment of numerals in this way limits the mathematical operations which can be validly performed on the measurements. As the divisions along the scale are not equal, addition is not a formally valid operation. An entity ranked 2 by an observer on a variable and ranked 3 on the same variable by another observer does not necessarily have a rank of 5 when the two observations are combined. Also if one entity improves its rank from 10 to 7 while another improves from 7 to 4, this in no way implies equal gains in the characteristic being measured. Further, multiplication is not formally possible. An entity ranked 4 does not necessarily have twice (or half) the amount of a characteristic as an entity ranked 8. Being denied these important mathematical operations implies that variables measured on ordinal scales only play a limited role in indicator development exercises. Their most important contribution is that they allow systems to be ranked in some sort of way. Generally, however, further analytical work can then be undertaken to extend or capitalize on these rankings.

Interval scales

Interval scales refine ordinal scales mainly by making the divisions along the scale represent equal quantities. The scale now does not only indicate relative position but it also indicates how large the difference is between two entities. The mathematical operation of addition is thus allowable. Interval scales only have an arbitrary zero point however and this is generally established to indicate a feature other than the absence of that quantity. A good example of an interval scale is the centigrade scale for measuring temperature. The zero point here does not indicate no temperature but simply the point at which water changes from its solid to its liquid state at normal atmospheric pressure. Numbers are assigned sequentially to divisions along the scale both above and below the zero point.

Ratio scales

A ratio scale is the most sophisticated of all four measurement scales discussed here. It possesses the properties of an interval scale but has the additional property of having its defined zero point represent absence of the variable. Almost all education system variables depending on a count of quantities – for example, total enrolments, numbers of buildings, financial outlays, etc – are measured on ratio scales. Data of this sort can be operated on mathematically by the operations of addition and multiplication. For example, an enrolment of 1000 students is twice that of 500 students while expenditure raised from Rp3 million to Rp4 million indicates the same absolute change in expenditure as a change from Rp30 million to Rp31 million.

The differences among the scales

The distinction between measurement of kind and measurement of degree is sometimes drawn to distinguish between types of indicators. Glock and Stark (1965) do this, for example, in discussing indicators of religious involvement and commitment (see especially around p71). They argue that in their context particularly, it is important to distinguish between the things people do (ie indicators of kind) and the number of people who do each of those things (ie indicators of degree). Separating these concepts can lead to more valid indicators of religious commitment. What these researchers are in essence claiming is that for constructing indicators, variables measured on nominal scales are just as important as variables measured on ordinal or more sophisticated measurement scales.

In education system measurement, the claim is often made that ordinal scales form a more adequate basis for measurement than do the other scales. This argument is generally based on the proposition that the magnitude of errors in data collection is generally unknown although it may be assumed to be fairly large. Hence very little faith can be placed on the absolute value of an observation or on the difference between values. More reliable estimates can instead be made of the relative order of a group of entities. Although this argument is partly true, it avoids two issues important for indicator development in particular.

Firstly, it avoids admitting that in most research applications, where both ordinal and a more sophisticated scale could be used, the results of analyses are very similar when even a moderate number of cases (say 15 or more) are used. This requires the distribution of the variables to be fairly symmetrical – as indeed they should be. When this is the case, parametric correlation coefficients and analyses of variance, for example, provide very similar results to their non-parametric counterparts. As data measured on interval or ratio scales allow more powerful mathematical operations to be used, adopting an ordinal scale for measurement can sometimes have quite limiting implications.

Secondly, advocating ordinal scales for indicator development precludes the need, which is often a very important need, to estimate changes in levels of educational provision. For example, it is not informative to a planner to know that more students are enrolled in special education courses in a given year than ever before. Instead the planner needs to know how many students there are now and how many more there are than previously. There is thus an implicit need for data measured on a ratio scale. As is indicated in Section 2.5, the error problem does not always pose the insurmountable problems often ascribed to it. Hence on these grounds, the development of education system indicators can and should proceed by using with confidence, variables measured of either interval or ratio scales.

There is a further corollary to the definition of these scales and the properties ascribed to them. Although certain mathematical operations cannot formally be applied to particular scales, this need not necessarily dictate the practical application of mathematics. Indeed, writers such as Lord (1953) and Baker *et al* (1966) have demonstrated that proper description and inference is possible by common mathematical manipulation of even nominal and ordinal scales. The implication from such articles is that ascertaining the scale on which a variable is measured does not have to be as crucial an issue as might be first thought.

These articles are very careful to point out, however, that very strict rules for the assignment of numerals to the divisions of the scales must be formulated and later adhered to when the results of calculations are interpreted. The implication of this corollary is that variables to be combined to form an indicator must allow the arithmetic operations to occur meaningfully. To ensure this, distinctions between the various types of scale must be carefully noted. Normal precautions would then safeguard against the definition of an indicator which required a variable measured on a nominal scale to be combined with a variable measured on a ratio scale.

The selection of a scale

In deciding on the sort of scale to be used to measure variables – and hence subsequently the sort of scale along which the values of an indicator will range – a balance must be effected between the precision of the available data and a need to adopt a scale with as much measurement sophistication as is possible. Any final data tabulation should use a scale reflecting this balance. For example, data measured directly with a ratio scale might be very inaccurate. To accommodate this, one way in which the data could be reported is to form class intervals of a width sufficient to exceed the estimated magnitude of the error. Hence intervals of ten or more units might be used on the assumption that this would allow most entities to be correctly classified. Such a modification would preserve the ratio scale – although it might be wasteful of the data. It would also conform to Scheuch's recommendation that 'the intervals for grouped data and the threshold of sensitivity for a measure should be greater than the possible errors in the data' (1966, 140). An alternative course of action would be to redefine the scale and merely rank the entities on an ordinal scale. This would then admit that observed differences between entities were artificial. Such a procedure can, on occasions, be most proper although it does destroy information contained in the data.

In other situations, the reverse of this problem arises. Variables are measured on a scale which does not extract the required degree of information about the characteristic being measured. For example, the level of literacy in a country might be rated as being 'high', 'medium' or 'low'. Such a measurement scale, although acknowledging errors which might be made in measuring literacy rates on other scales, is a crude approach. It does not allow very many distinctions to be made amongst the countries being examined. Instead of an ordinal scale such as this, a ratio scale measuring the percentage of the population who are literate (according to some definition) could be estimated. Errors here would be certainly greater and hence the reliability of the observations would be less. However, using such a scale would allow decomposition of the scale to reduce the effects of the error at a later time – even perhaps to the three-point ordinal scale. Banks and Textor (1963) used this 'change of scale' approach in their *Cross-Polity Survey*. They reduced all of the scales on which their variables were measured to two-point ordinal scales – even though many of the variables had a high degree of reliability. The reduction meant that Banks and Textor lost many very valuable distinctions between nations. The only explanation they offer for doing this is that it 'simplified the analysis'. With computer methods available, this is not an adequate explanation by itself for using an unsophisticated scale – especially in view of Ebel's (1969) demonstration of how considerable the reduction is in the consistency (reliability) of the measurement when two-point rather than four- or five-point scales are used.

What scale is proper and required in a particular situation must be ascertained very carefully. Preference should be given to those variables having the highest level of information and measurement on the most sophisticated scale. This would then enable more powerful but still reliable and valid indicator development to take place.

2.3 Specific features of education systems data

The previous sections have described certain similarities and differences between the data base for variables involved with the measurement of education systems and the base involved with the more common psychological and educational assessments. In addition to these differences, certain other features of education systems data should also be discussed.

Many education systems data are indirect measurements

One difference between education systems data and psychological data is that education systems data are mainly collected for purposes other than for use in planning and research in general or for indicator formation in particular. Some of the many data generated from our societal sub-systems form the basis for social welfare payments; others may act as a check on the control of infectious diseases while still others are used to allocate students, teachers, equipment, finance, etc to schools or other educational institutions. As the various sub-systems become increasingly more complex, the generation of data is of utmost importance for a variety of accounting purposes.

As a by-product of these procedures, data tabulations are often produced for wider circulation. These tabulations are aggregated or sub-divided in ways which are commonly considered to be useful for reporting. National and state (or provincial) yearbooks contain many tables of this type. It is in turn from tabulations such as these that a great many of the data on which indicator development must be based are derived.

Such data are called 'indirect' or 'secondary' measures. The purpose for which they are used in this context is different from the purpose motivating their original derivation. Opposed to these kinds of data are 'direct' or 'primary' measures – the intended use of these dictates the methods of data collection. Direct measures, for example, often come from special purpose surveys and interviews.

The implications of having to use indirect measures for indicator development are numerous. Perhaps the major implication is that if data collected primarily for other purposes were not used, very little data would be available. Depending on the sort of data and the system it comes from, many difficulties would prevent collection. The greatest difficulties – and two which could not be overcome – are the allied ones of costs and time. It would imply an undertaking amounting to an annual census of proportions never before considered. Indeed the concept of annually asking every person in a country whether, in the previous year, they had been unemployed, eligible for social security payments, what diseases they had had, what crimes they had been convicted of, what education or training they had been engaged in and what their gross income had been, etc is horrendous and the task impossible. Censuses even on the present scale, are very expensive undertakings in terms of costs, manpower and time. Hence data collected from a system for specific purposes must be used for other purposes provided that those data approximate the concept defined as being of concern.

Another implication is that the precise concept as described by the definition of a variable or indicator is not necessarily measured in the way required. This problem is known as 'fractional measurement'. As the main effect of this problem is to reduce the internal validity of the measures, further consideration of this implication is deferred until the next section. It is sometimes possible to modify the process of data collection to ensure that a slightly different data base is obtained or that a disaggregation of the data can occur. When this happens there exists a strong inter-relationship between the purposes for both the direct and the indirect measurements. Normally, however, such modifications do not occur.

The nature of the data required for indicator development really implies that heavy reliance must be placed on data of an indirect nature. This does not necessarily weaken an indicator. Sometimes a strong correspondence exists between what is required for an indicator and what is available for use. The extent to which indirect data do weaken an indicator must however be ascertained in each situation. Some implications for indicator development of those instances where a strong correspondence does not exist are discussed in the sections on reliability and validity.

Many sources of indirect measurement exist. Ministerial reports, reports from statistical authorities and reports of large companies all provide a variety of indirect data. The reliability and validity of these data are, however, unknown. For education systems within a country, the extent to which data are checked for accuracy and their correspondence to established definitions can be ascertained fairly easily. The comparability between areas and/or between years for data on particular defined variables can then be estimated and cautions sounded as necessary.

Sources of data on education systems in a number of countries are principally the various UN *Yearbooks* as well as the *Yearbooks* from the UN agencies such as Unesco and UNICEF. Also to be included here are the OECD reports as well as World Bank and Ford Foundation reports.

Estimating the extent of comparability of international data can be an extremely difficult undertaking. At the end of some tabulations in reports issued by the above agencies, a series of cautionary notes is normally provided to warn how the entries for some countries include certain aspects of those systems while the corresponding entries for other countries exclude those aspects. Differences in definition of the concepts are also noted. Recently, in order to overcome some of the difficulty in reporting educational statistics from its member countries, the OECD undertook a programme aimed at achieving maximum practical comparability in its reporting. The results of this undertaking have been published as a series of ten reports titled 'Classification of Educational Systems'. On a much wider scale, Unesco has published the *International Standard Classification of Education*. It can only be hoped that efforts along these lines are continued and diversified to gain greater comparability of inter-education system statistics across as many countries as is possible.

Another problem of using indirect measures is that of selecting a particular source for reference. This is an important problem which is often not even considered. Even though a number of sources might purport to publish data on the same variable, the values recorded in their tabulations are not necessarily the same. Table 2.1 reports some differences selected at random between UN and Unesco *Statistical Yearbooks* while Table 2.2 reports some differences between statistics for the first- and second-levels of Australian education as reported by Australian sources and the corresponding data as reported by Unesco. The point of these tables is to demonstrate that sources of indirect data can differ considerably – sometimes inexplicably – from one to another. The most authoritative source for reference for a particular type of data should therefore be determined and consulted consistently when compiling data tabulations for indicator work. Changing from one source to another can often lead to difficulties which can then erroneously reflect on a country's level of educational development or on its development efforts rather than on the variety and differences in data sources.

In addition to the reports of the international agencies, a number of authoritative handbooks have also been published as well as a number of data banks established. Amongst the former are those by Banks and Textor (1963), Russett *et al* (1964) and Taylor and Hudson (1972). Amongst the latter, and sometimes associated with the former, are the Yale Political Data Program at Yale University and the Dimensionality of Nations project at Northwestern University.

The units of education systems are not always comparable

A second feature of education systems data is that the comparability of the units and the systems being analysed must be ascertained before any analysis can be undertaken. The necessity for having to do this stems from the ways variables are defined as well as from the ways in which data are collected. The difficulty noted here is that territorial units, although being clearly defined, might be subject to internal or external influences which undermine nominal comparability. For

Variable	Country	Year	UN estimate		Unesco estimate	
			Value	Source*	Value	Source†
1st-level students	Afghanistan	1971	604795	1974 (821)	572935	1974 (176)
1st-level students	El Salvador	1970	509985	1974 (826)	531309	1974 (173)
3rd-level students	Gabon	1970	65	1974 (827)	172	1974 (304)
3rd-level students	Peru	1971	135351	1974 (835)	129472	1974 (318)
% of GNP on education	Canada	1969	8.9	1971 (776)	8.3	1971 (511)
% of budget on education	Guatemala	1969	18.6	1971 (777)	17.5	1971 (513)
% of budget on education	Brazil	1969	17.5	1971 (776)	15.2	1971 (513)

* Year of the UN *Statistical Yearbook* and page number
† Year of the Unesco *Statistical Yearbook* and page number

Table 2.1: *Some differences in values of UN and Unesco reported data*

Year	Enrolments (000s)		%
	Australian government	Unesco	Difference
1962	2,253	2,361	4.79
1963	2,309	2,421	4.85
1964	2,367	2,465	4.14
1965	2,438	2,576	5.66
1966	2,504	2,561	2.28
1967	2,586	2,742	6.03
1968	2,656	2,849	7.27
1969	2,717	2,892	6.44
1970	2,768	2,944	6.36
1971	2,808	2,940	4.70
1972	2,841	3,022	6.37

Source: *Schools* (various years) and Unesco *Statistical Yearbook* (various years)

Table 2.2: *Australian first-level enrolments as reported by the Australian government and Unesco*

example, a newly independent country might be able to afford quite substantial expenditures for education despite its lack of resources. In such cases, a bilateral agreement with the previous colonial power might be used for a number of years to ensure that level of educational funding remains high. Similarly a country subjected to internal conflict and insurgency might not actively encourage second- and third-level enrolments but instead recruit young people for the defence forces.

Comparing countries in either of these situations at one point in time with other countries not subject to these influences is thus not a proper comparison. The collectives are not comparable. In neither case can proper assessments be made of the educational development level nor of the priority a government may wish education to have. Comparing nations with different political ideologies can present similar problems in cross-national studies.

The alleged distinction between quantity and quality

In discussing scales used in indicator development, the question must inevitably arise as to whether the use of measurement precludes those aspects of education systems which are important quality concerns. Many arguments are proposed to support the view that quality considerations of an education system are much more important for planners and researchers than 'mere quantitative' considerations. Such an argument, despite its widespread propagation, is not considered valid in this context of indicator development.

In common (not statistical) usage, the term 'quantitative' is considered to reflect only the amount of a variable present in a given situation. The term 'qualitative' is considered to identify the relative excellence of that amount as though the two aspects can be distinguished from each other. Sometimes indeed qualitative considerations are considered to have the essential property of being non-quantitative. As Coombs (1969, 15) points out however, the alleged distinction is artificial. A statistic can be operationally defined so that its use and interpretation will classify it as being quantitative or qualitative or both. For example, McGranahan *et al* posit that the quantitative variable 'enrolment ratio' is often criticized 'because it does not reflect quality of education. It is quite possible, however, that by and large, countries with higher enrolments also have better equipment, better quality teachers, fewer drop-outs, etc, so that the enrolment ratio may, in fact, reflect quality indirectly as well as quantity directly' (1972, 52).

Eide attempts to explain why the artificial separation of measures of alleged quality from measures of alleged quantity occurs. He notes that 'the frequently applied distinction between "quantitative" and "qualitative" aspects is mainly a convention. Largely, such distinctions reflect previous priorities in terms of the development of measures and actual data gathering. We have usually been able to quantify what we have found important enough to be concerned with' (1973, 6).

An example demonstrating the inseparability of measures of quantity from measures of quality is the Consumer Price Index. This indicator combines estimates of the quantity of goods bought by a family in a certain period with the costs of those goods. It could therefore be regarded as being a quantitative indicator. The 'basket of goods' on which the calculation is based, however, is carefully chosen so as to indicate a certain level of quality of life. Indeed, it is also updated from time to time to reflect a change in the quality of life of citizens within an area or a nation. Such changes involve selecting similarly labelled products which are of a different level of quality. The CPI therefore contains both qualitative and quantitative dimensions.

All qualitative considerations of an education system can be operationally defined and measured on at least a nominal scale. Although this might be a very primitive and basic measurement, it nevertheless does allow a quantification of qualitative concerns. It is highly likely that many variables could be defined in this way and that an indicator of quality could combine a number of these simply measured variables. Such an approach does facilitate the inclusion of measures reflecting those features of an education system considered to be qualitative in kind. In addition, other variables measured on one of the more sophisticated scales can certainly be defined in some cases. This thus underlines Philp's assertion that 'quality can be defined and measured in meaningful and precise terms for purposes of analysis and measurement' (1969, 280). For these reasons, an explicit distinction between indicators (or variables) measuring quantitative concerns and those of a qualitative nature is not made in the present discussion.

2.4 Internal validity and operational definitions

The concept of internal validity is explained by Etzioni and Lehman as being 'the extent of correspondence between a social science concept and its operational definition' (1969, 46). It is therefore concerned with that part of the measurement process during which the variable actually being measured is compared to what was intended from a theoretical or conceptual standpoint. Almost inevitably there must be some difference between the desired and the actual measurement. When this gap becomes too large, and what is measured bears little relationship to the original concept, the variable – although possibly measured consistently – loses its usefulness. It has then very little internal validity.

Operational definition of a concept into indicators or variables occurs in all science. It is an attempt to state precisely, for purposes of measurement, what will represent a more general concept. This practice is required because the concept itself is normally not directly measurable. The defined indicator or variable must thus act as a surrogate for the concept.

Three main influences can reduce internal validity of education system indicators and variables. The first of these is fractional measurement. This occurs when a number of different aspects of a concept are identifiable but only one or two of these aspects are actually measured. A fraction of the concept is thus represented. The aspects of the original concept not measured represent the reduction in internal validity. The concept of female participation in education might be defined, for example. This could be represented by a number of different variables including the percentage of pupils or teachers who are female at the first-, second- and third-levels of the system. Selecting only one or two of these variables gives a very biassed impression of female participation in a particular system by only fractionally representing the concept defined.

As Etzioni and Lehman point out, fractional measurement must reduce internal validity of most concepts because 'it is extremely difficult to construct an operational definition that covers a concept in all its ramifications and dimensions' (1969, 46). Attempts must therefore be made at least to minimize the effects of fractional measurement on internal validity.

Indicators, as defined in Chapter 1, are one way in which the internal validity of measures of education systems can be increased. Variables taken individually – as shown in Table 1.1 – can only measure one aspect of a concept containing many equally important aspects of educational provision. An indicator might therefore be expected to have higher internal validity than a single variable. Simply forming an indicator is not, however, sufficient justification to allow this claim to be made. The several variables representing each aspect of a concept must be identified, reliably measured and then properly combined to form the composite indicator. If only some of the main aspects of a concept are represented or aspects more properly reflecting another concept are included, internal validity must be significantly reduced.

A second influence reducing internal validity of measurement is variability of concept. This is manifest by a concept changing in its implications either over time or between education systems although the operational definition of that concept is not necessarily different. Cross-national comparability of definition has already been discussed in previous sections. It is a difficulty which is also significant in this context. An example of the difficulty implied is provided by an attempt to define the concept of secondary education. In developed countries, a number of years of this form of education is compulsory – at least in practice if not by law. Without a certificate of education attainment from this level, few economic opportunities are available. In a developing country, however, even a little secondary education can have a significant effect on an individual's earning potential. Government and commercial positions and teaching posts are immediately available, for example. Comparing data on secondary enrolments for very different sorts of countries can therefore be difficult because of the different implications secondary education has in those different contexts. This difficulty is also related to the different objectives held by the politicians and planners in the different countries. Within levels of education, different kinds of education or training are also given. Some systems have very elitist types of school (eg the grammar school in England or the Berufsschule of Germany) while other systems provide more general education with equal opportunity. Further, at the tertiary level, what amounts to a high level graduate from a surgery or physician programme in one country may be equivalent to a second or third year undergraduate medical student in another country. Such differences as these in concept severely reduce the comparability of data from different education systems and this is thus a reduction in the internal validity of those measures.

Differences in variable definition can also be an obstacle to analysis and interpretation in many situations. Bergson (1953) and Campbell (1959) both demonstrate how Soviet economic indicators have been defined and the implications these definitions have for purposes of comparison both with other countries as well as with Soviet indicators for other periods. Orleans reviews the traditional approach to statistics in China. 'The Chinese measure of area, the mou, varies not only from region to region, but also over time . . . at present the mou is equal to 0.1647 of an acre, but over the past centuries it has fluctuated from between 0.1133 and 0.1669 acres per mou' (1974, 47). Orleans then describes the lack of statistical control over the collection of economic statistics as well as the adjustments many Party officials feel compelled to make: 'As statistics move up to people with more and more statistical sophistication, they are subjected to more and more "corrections". With the advantage of having past records available to them, personnel up the line feel they are in a better position to know "what should be" while at the same time meeting the criteria of what may be expected' (1974, 51). Such changes to definition of variables make interpretation and comparability of statistics impossible.

A third influence reducing internal validity is the definition of a variable to represent a concept even though it is known that the variable does not represent that concept properly. Such a practice can be termed 'concept substitution'. For example, an attempt might be made to measure the quality of the output of an education system. If by quality is inferred achievement level, the way measurement might be attempted is through variables defined as the achievement test scores for graduates at various levels. As such scores are generally not available, graduation rates might be defined instead. These rates then act as a substitute for the achievement levels. As there is no necessary relationship between these two concepts, the substitution must change the implication of any conclusions drawn. This subsequent modification to the results does not occur very often in research studies, however.

To minimize the problem of having to decrease the internal validity of measurements, one tempting solution is to define a concept by its operational definition. The logical order of the procedure is thus reversed. This is a commonly used practice in psychological testing, although in that context it can be quite an appropriate one. A definition of a concept can be made in terms of the test used to measure that concept. For example, need-achievement might be defined as being whatever is measured by a particular test. Such a procedure as this, unfortunately, can often be more implicit than explicit. For education systems, this form of procedure is probably also more common than is suspected. Expenditure on education is defined as being the sum reported in official documents and any components of that sum are thus valid components. Expenditures incurred on behalf of other activities also included in the sum or educational expenditures attributed to other budgets both serve partly to invalidate the original definition. Such a practice of reverse definition is not to be commended for systems data. Etzioni and Lehman very strongly urge alternative action to overcome this problem: 'To act as if an operational definition were automatically the same as the underlying concept is a questionable procedure, and it is also likely to have important negative consequences in the realm of policy-making' (1969, 54).

Because differences exist in both intra-national and international data tabulations, the major practical implication for determining the internal validity of data is that differences in definition used to form the data tabulations must be noted. When major discrepancies occur, attempts might be made to identify those components which are comparable. Less powerful analyses might be attempted and less mathematically demanding measurement scales might be adopted.

In addition, the attempts by the international agencies to gain comparability of data from different systems and discussed earlier must be supported and encouraged. If more systems adopt the commonly accepted definitions and concepts, the greater is the scope for both inter-system as well as intra-system cross-temporal comparison.

The internal validity of an indicator must be assessed to be high – at least high enough for the specific situation for which that indicator is to be used. Unfortunately there is no sound statistical way in which the congruence between an operational definition and its concept can be quantified.[1] This then raises the problem that a slightly different operational definition of a concept – which might appear to be just as valid – could produce slightly different relationships or classifications. As Lazarsfeld concludes, this problem 'is unavoidable because indicators at least have only an inferential relationship to the underlying factor sought' (1959, 61). In any particular situation, therefore, when two or more operational definitions can be derived from a concept and their measurement properties are equal, the most appropriate definition for the situation must be selected as the one best defining that indicator. Sometimes the method of interchangeability of indicators, as described in Section 3.5, might also be used.

---

[1] A possible non-statistical way would involve asking a panel of 'judges' to give their 'expert assessment' on the congruence between the operational definition and the concept.

2.5 Reliability and the consistency of measurement

Ascertaining the reliability of data is also a central issue in any measurement operation. It is concerned with determining whether data can be claimed to measure (whatever they do measure) with consistency. If data are reliable, the same results could be consistently obtained if the same characteristics were measured again by different individuals. If two different sets of data were obtained from a system for the same defined set of education system variables at the same point in time, the data in both sets would be classified as being unreliable. The hallmark of reliability is thus the extent to which a measurement operation can consistently produce similar results.

Few variables are ever measured perfectly reliably. Hence education system data reported in tabular form – and subsequently used for indicator development – must be acknowledged as not representing the real situation exactly. A difference exists between what may be termed the 'true' education system and the 'observed' education system. This difference is brought about by a number of intervening influences, each of which can induce error into the measurement. These errors accumulate in an unknown way so that it is generally impossible to determine how large the errors are. Diagram 2.1 shows schematically that there are at least four sources[1] of measurement error:

☐ the consistency of applying an operational definition
☐ the data collection procedures
☐ the data collection instruments
☐ the data processing procedures.

The data collected from the true or real education system is filtered through each of these potential sources of error. The cumulative effect of these four filters is incorporated into the education system statistics finally reported. If the effects are not too large, the reported system should reflect the relative magnitude of variables in the real system fairly closely. As it is impossible to tell how close the match is between the observed and the true system, considerable effort needs to be directed to making the data collection stages as efficient and as effective as possible. By doing this, it can be anticipated (although not guaranteed) that the data have a fairly high degree of reliability.

If a set of data is reliable, no implication can be made that the characteristics claimed to have been measured actually have been. Hence quantities reported as being present in the observed education system and their distribution may or may not reflect the situation in the real education system. To impute this congruence is possible only after determining a high level of internal validity for a data set. A high

---

[1] Actual changes in the system itself are not, of course, considered here. The measurement operation being discussed is assumed to be conducted with the aim of assessing the situation as it exists.

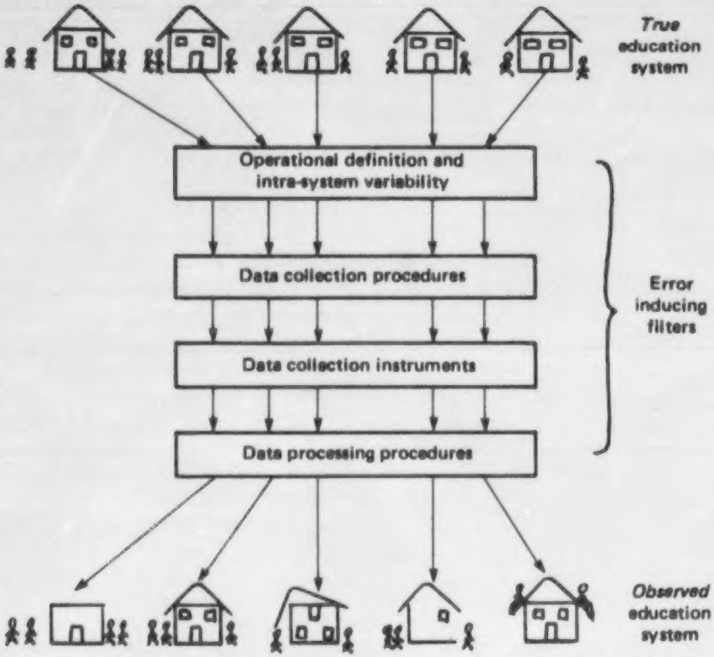

True education system
Operational definition and intra-system variability
Data collection procedures
Data collection instruments
Data processing procedures
Error inducing filters
Observed education system

Diagram 2.1: *Sources contributing to reducing the reliability of education system variables*

reliability only implies that something has been measured and that the measurement operation is such as to produce consistent results.

The consistency of applying an operational definition

The first way errors can be induced into data is through not using the operational definition of a variable in a consistent manner in all the units from which data are collected. Partly this may be due to the definition being ambiguous or incomplete. Ambiguity, for example, allows a definition to be interpreted in some ways by some individuals and in other ways by other individuals. Such differences in interpretation can occur within schools, within local planning offices or within national ministries. The ambiguity thus leads to inconsistency because equivalent pieces of information are being treated in different ways in different places.

The variable 'enrolments at the first-level' is one which often causes confusion. It is exceptionally difficult to compare data on this variable (or any of its derivatives – say the first-level enrolment ratio) on a cross-national basis. Without defining the variable very explicitly, confusion immediately arises as to what constitutes education at the first-level. Are, for example, special education classes, physically attached to but still administratively independent from primary schools, to be included? Are pupils attending pre-school or kindergarten classes to be counted as being enrolled at the first-level when only some of these classes are attached both physically and administratively to primary schools? In some countries decisions related to what is and what is not to be included are clear. In other countries, and even between some provinces or states within a country, ambiguity arises about the principles determining inclusion. Generally an official decision can clarify the operational definition to avoid the ambiguity and hence obtain consistency in the data collected. When these decisions are taken at the state level only, however, and each state within a country acts independently of others, national statistics then become unreliable. This is compounded further when many countries are being compared. Hence, for indicator development, data should be derived from sources considered to report reliable data.

Even with very clear and concise operational definitions, education systems still vary according to the time at which the data are collected. Enrolments at the beginning of a year always tend to be higher than those observed about the middle of the year, while those at the end of the year tend to be still lower. Differences in the months of commencing the school year – especially between countries in the northern hemisphere compared to those in the south – thus introduce another source of unreliability. This problem of course does not generally occur with data being considered for just one country – unless the date on which annual statistics are reported is changed.

Another way in which an operational definition might not be consistently applied is in a situation where certain individuals responsible for providing data take it upon themselves to redefine the variables. Such people might consider, for example, that a high enrolment figure for a school for a particular area in which a government has been attempting to reduce illiteracy, might be favourable to the government. (The reverse attitude might, of course, be adopted instead.) Data might therefore be artificially inflated (or deflated) so as to communicate the existence of a favourable situation. Implicit redefinition might also occur if enrolment figures are not considered to be at all important by certain individuals. If no check is ever made of the data provided to a central office – as is so often the case – then a teacher or ministry official may be content to report the same set of values for a number of consecutive years – regardless of the congruence between those values and the real situation. Such a practice has certainly been identified in another area of national accounting (see, for example, Dewey, 1974). It can also happen that when a survey is deemed to be important by individuals, instructions and definitions are generally very carefully followed. When this occurs, reliability is increased.

Errors induced by the data collection procedures

The second way errors are introduced into a data tabulation is through the procedures employed to collect the data. For most types of education system variables – and especially those most frequently referred to – data should be collected either as part of a complete census of the education system or from a sample survey. Both of these methods have their advantages but they also have their attendant disadvantages.

Census methods for collecting data are employed when a complete enumeration is required. In the context of an education system, this is generally a feasible undertaking on an annual basis in one of a number of ways. These may range from collecting information on each individual within a system to collecting information on aggregates within grades.

An attempt to complete data coverage is normally undertaken in conjunction with other administrative procedures such as school classification, certification, payment of subsidies, etc. For these reasons, the extra cost of the census over a sample survey can often be justified. When such large amounts of data are collected, however, errors at the individual person or individual institutional level due to non-reporting, partial reporting and/or incorrect reporting occur. These errors are often impossible to overcome – or even to detect.

In contrast, data derived from samples do not aim at a complete enumeration. They instead attempt to estimate the characteristics of the population from a smaller number of respondents. The more care taken in selecting the respondents included in a sample and in obtaining their responses, the greater the faith which can be placed in generalizations from the sample characteristics to those of the population. Numbers in a carefully selected sample do not have to be very large and hence the volume of data obtained from a sample is very much more manageable than that derived from a census.

Variables which might be incorporated into education system indicators do not necessarily have to be of the type normally now asked. The conceptual distinctions raised in Diagram 1.1 demonstrate that very many variables not presently collected should also be considered for possible combination into useful indicators. Collecting data on variables not anticipated by respondents in the current annual procedures can be best effected by sample survey procedures employing questionnaires and/or interviews.

In such circumstances as these, errors can be induced in the responses in various ways. Dewey (1974) is one who investigates this problem. He reports that the educational level of a person charged with collecting statistics can influence the magnitude of the error. Field workers with low levels of education or people who are not trained to appreciate the need to collect good data and to recognize the difficulties which can prevent data being reliably reported will contribute errors of an unknown magnitude to any final data collation. Schuman and Converse

(1971) show that other personal characteristics of an interviewer can also affect the magnitude of response error in certain types of questions.

Many considerations affect the selection of a census or a sample as the method of obtaining data from an education system. On the basis of empirical evidence, one fact is clear – complete enumeration can be far less accurate than taking a sample (see, for example, Deming [1950, 358] and Dewey [1974, 9-10]). In addition, sample surveys are less expensive in terms of costs and time. Hence other considerations being equal, sample survey procedure should be preferred over census procedures as a data collection method for many education system data.

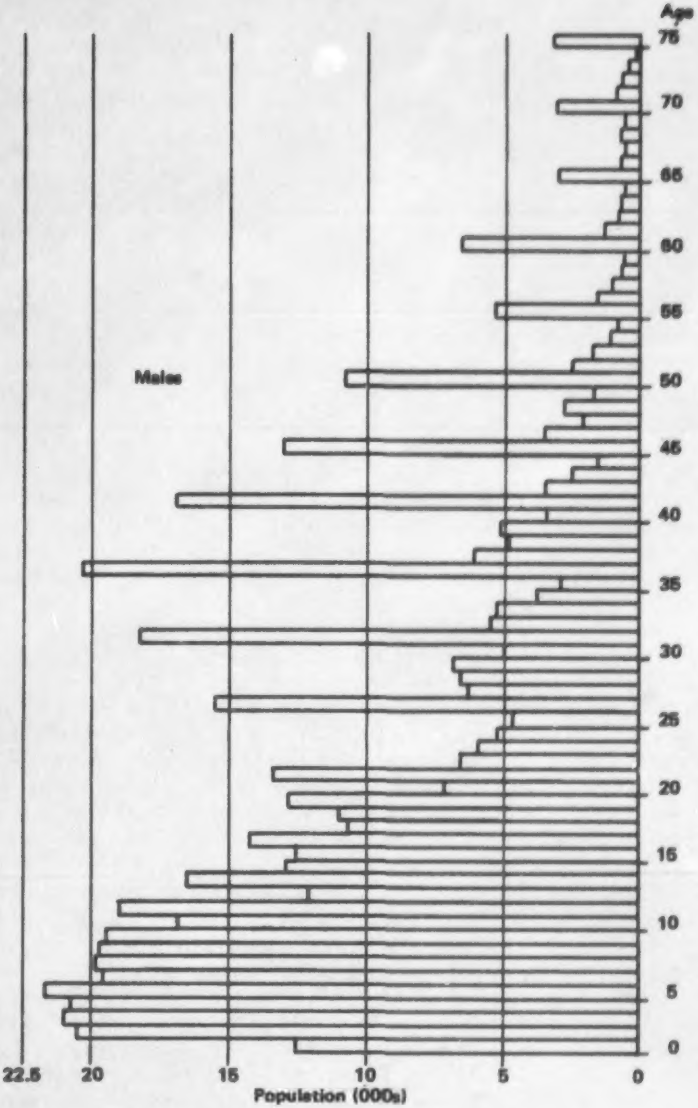
As data can be collected in different ways, it is essential to consider the implications for each variable of alternative methods. Different procedures induce different degrees of error. Hence those procedures tending to induce least error or which are less susceptible to bias-inducing error should be identified and adopted in preference to other methods. An important criterion in selecting an appropriate data gathering procedure should therefore be its susceptibility to the introduction of error into results.

Errors induced by the data collection instruments

Errors contributing to unreliability of data can also be introduced in yet a third way. This is through the instrument used for data collection within the type of procedure adopted. Each type of instrument has its own peculiar difficulties. Many of these are elaborated elsewhere (Johnstone, 1978b).

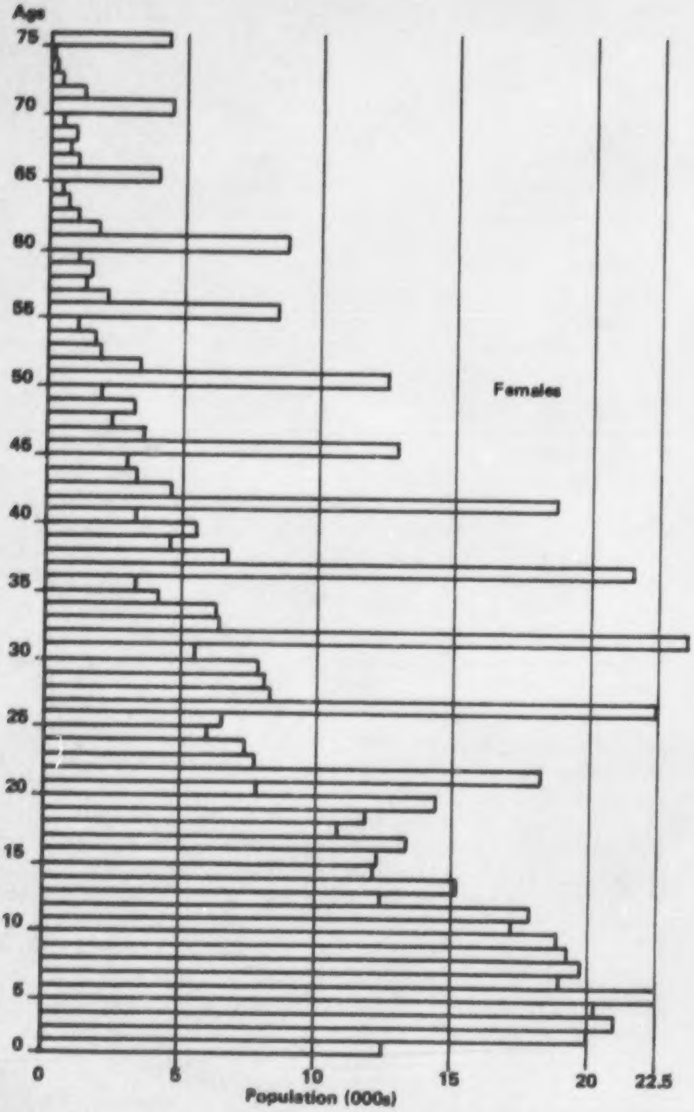
Annual returns and questionnaires, for example, can be confusing in the sort of data they require. If instructions do not precisely specify or define the variables on which data are required, a quantity might be estimated on the assumption (or hope?) that the value provided is possibly what is wanted. Unfortunately, imprecision of this kind can often happen in schools where the person in charge is very busy because he is more concerned with the process of education in the classroom than with the recording of data. In such schools it can sometimes be that more useful data are available than those provided from other schools in which the business of education is more pedestrian. It is essential, therefore, that annual returns especially are designed to be clear, quick to complete and unambiguous.

In other situations, errors can be induced simply because of the nature of the variable on which data are being collected. Population data frequently underestimate the true situation in a country – particularly in large countries. Similarly age data underestimate some ages while, at the same time, overestimating other ages. Diagram 2.2 demonstrates this imbalance in the case of Indonesia although the majority of developing countries also manifest the same phenomenon. The higher frequencies of people with ages ending with a zero or a five digit (eg 20, 25, 30, 35, etc) compared to the frequencies of those with



Source: Biro Pusat Statistik (1973), p13

Diagram 2.2: *Population pyramid for Indonesia – 1971*


Age
75
70
65
60
55
50
45
40
35
30
25
20
15
10
5
0
Females
0
5
10
15
20
22.5
Population (000s)

ages ending with another digit are not realistic. Age to the nearest five years, rather than an exact answer, is thus often given in reply to questions about age.

Surveys and interviews can contribute errors of measurement through the nature of their construction and application. In the first of these ways, both types of instrument can have similar difficulties in ambiguity as does an annual return. In the second way, they can induce error through the differential reinforcement or provision of information by various field workers. Consider, for example, a series of questions in an interview aimed at determining both the gross and the net income of a teacher as well as the sources of income (eg salary, honorarium, gratuity, school fees, etc). A field worker not fully conversant with the situation or who did not speak fluently the dialect of the person being interviewed, might tend to accept any answers provided without requiring elaboration or clarification. Another field worker, however, who felt confident in both respects might question and even point out apparent contradictions in the answers provided. These contradictions could then be discussed with the respondent with the possibility of making a change to the original answers in the light of reconsideration. Comparing the responses from two interviewers such as these would demonstrate considerable unreliability in the collected data. Shapiro demonstrates how an interviewer can goad a respondent 'into mentioning more items for each of (a set of) open-ended questions while others were, in varying degrees, less aggressive, thereby eliciting fewer items' (1970, 414). Collins (1970) claims this form of error could be identified and accounted for because a single interviewer tends to induce a fairly consistent pattern of responding across all interviews conducted by him.

Achievement tests used to measure some resource outputs of an education system very often have quite low reliability. These reliability values also vary according to certain background characteristics of the examinees. The socio-economic level of the examinees, their colour and their ability are three such characteristics. These influences further complicate inferences which might be made from observed scores.

Errors induced by instrument construction are generally well recognized. They tend to be the one source of error attention is directed towards in data collection. This then tends to minimize the error attributable to this factor. Such recognition does not, however, ensure that no error is induced from this source nor that it is no longer an important consideration. It is instead a stage in which many technical advances have occurred.

Errors induced by the data processing procedures

The fourth source of error depicted in Diagram 2.1 is the data processing procedures developed to collate and process the data. Some of the error induced here can be termed 'systematic error' because it can stem from a deliberate attempt by the processing officials to over-estimate or under-estimate values. An example which demonstrates this is where Schwarz alleges that 'precise facts and figures are scarcer in Liberia than elsewhere: lack of economic development in the interior and an unofficial feeling that facts can be dangerous have combined to make them so' (1969, 296).

Other errors are not induced so deliberately nor consistently. They are instead distributed in such a way as to allow them to be termed 'random errors'. Simple errors in transcription are included here, as too are the accidental omission or double counting of some returns. (For a discussion on this point see Rummel [1970a] and Russett [1965, 444].) Indeed both systematic and random errors have been alleged to be country specific. When commenting on the processing of data, for example, Coleman alleges that:

> in one study, it was said the key punching errors reflected differences in national character: Germany showed a total absence of random errors in key punching; whenever an error occurred it was a systematic error that appeared in all cards; in Italy the errors were scattered with apparent Latin abandon. (1969, 98)

When processing methods induce additional errors into a measurement operation, the methods of collection and checking for accuracy must be reviewed. With the advent of individual data systems and the high level of sophistication developed in the computerized processing of returns, errors at this stage should not reduce the reliability of data significantly if at all. When errors are introduced, however, quite bizarre changes can sometimes occur. Coale and Stephan note that the 1950 Census figures show 'a surprising number (1670) of widowed fourteen year old boys' (1962, 338). They attribute this observation to card punching errors involving the punching of data onto computer cards one column to the right of where it should have been punched. This has a cumulative effect on many results. For example:

> A white person when relationship to the head was child would be coded as a male Indian, while a Negro child of the household head would be coded as a female Indian. If the white child were male, he would appear as an Indian in his teens; if female, as an Indian in her twenties. (1962, 345)

In many instances this error is detected, but in other cases it is not. Hence, random error is introduced through the data processing. The effect of the above errors on the 1950 US Census results is, according to Coale and Stephan, 'completely negligible' overall yet still noticeable.

Data processing operations are very much technical-practical ones in which very deliberate actions can be taken to maintain reliability at a high level, particularly if the other stages have been adequately supervised. It is not, however, a stage where errors introduced in other stages can be eliminated.

Two recommendations can be made on the basis of the foregoing discussion on reliability. One recommendation is that efforts must be made to ensure that data within countries are as reliable as is possible. Where local planning agencies and/or state departments of education exist, their data collection procedures should be co-ordinated to ensure practical similarity in definitions, procedures, instrumentation and processing. Special purpose investigations, when conducted, should also be required to contribute to the established data banks rather than, as so often happens, be independent and even unidentifiable. Individual countries can do a great deal to ensure reliability of their own data partly because they control their own destinies and partly because they provide a better basis than is possible for cross-national efforts. As Retzlaff posits, it makes 'good sense to devote some effort to the intra-national comparison work where data comparability both within and across time periods can be assumed to be of a much higher order of reliability' (1965, 77).

A second recommendation is that research be conducted into the significance of errors in education systems data. Partly this research should be aimed at identifying the implications of errors for indicator development but it would also identify implications about the planning process itself. The above discussion has had to be based on the tenet that errors are induced in various ways. No imputation has been possible about the likely magnitude of these errors. Even if this had been possible, the significance of a 'standard error' could not have been estimated. This is a real problem. It occurs also in the context of predicting enrolments, teacher numbers, etc by mathematical models. How much tolerance, for example, is allowable in estimating student enrolments before teacher recruitment and building construction programmes are significantly affected?

This problem manifests itself in all areas of social accounting. Indian agriculture statistics, for example, were extremely unreliable up until about the middle of this century. They were so bad in fact, that they were 'useless for working out whether or not food was increasing in proportion to population . . . . As the Bengal famine showed [in 1944], they were unfit even to warn the government of the advent of famine and when food rationing was introduced, food controllers complained bitterly of their [the statistics] inadequacy as a basis for food planning' (Dewey, 1974, 10). An analogous situation could arise in areas where estimates need to be made of the stock of educated manpower and the distribution of skills throughout certain communities in many underdeveloped countries. Conclusions and recommendations which were most inappropriate could be reached on the basis of wrongly reported or estimated statistics. It is important, therefore, to identify the nature of the various sorts of error. Their effects on the values calculated for indicators will then also be determinable. Until that stage, only single numbers can be reported for indicators rather than the error bands which are commonly reported for conventional test scores.

To this point, the discussion on reliability has concentrated on the reliability of single variables because these are what are initially collected. To the extent that errors are contained in variables, they must also be contained in indicators formed by combining a number of variables. With indicators, however, there can be some compensation for these errors in some situations. Classical test theory hypothesizes that as the number of observations increases, so too does reliability. As indicators combine a number of variables, the reliability of measurement of a concept also increases. Hence the use of indicators to comment on education systems should provide a more reliable basis than using a set of independent variables. Brandis (1970) is one who recognizes this possibility in forming his social class indicator, for example.

2.6 External validity and the independence of concepts

A final issue in the measurement of education system variables is that of ensuring external validity. This property, as defined by Etzioni and Lehman, is 'the truth of hypotheses that relate operationally defined categories to one another' (1969, 46). It is therefore a property of measurement concerning the way in which concepts inter-relate. It is a consideration which can only be made after the internal validity of an operational definition is assured and the reliability of a measurement has been ascertained or accepted.

The property of external validity implies that a concept should not contain elements defined as belonging to another concept. There should therefore be an independence of concepts. For definition of indicators, this requires that, in general, no variable contributing to one index should also be defined as contributing to another. Expressed in different terms, any definition of an education system should be parsimonious.

An example relating to variables demonstrates this concept. The following closely related input variables are all commonly cited in the literature:

☐ unadjusted enrolment ratio for the first-level
☐ adjusted (net) enrolment ratio for the first-level
☐ first-level enrolments per 10,000 population.

Each of these variables measures slightly different quantities but all basically measure the extent of enrolments at the first-level. To include the three variables in a data set would duplicate, in part, this information. To convey it more parsimoniously requires either that only one of the variables be chosen (ie form a representative index) or that the three be combined into a single index (ie form a composite index). Either of these approaches ensures that no single concept in the education system space is over-represented.

Also implicit in the concept of external validity is the principle that a set of indicators possessing the property should define the education system space completely. In research work this is a very commonly cited principle guiding the selection of variables especially for cross-national study (eg Cattell [1949], Adelman and Morris [1967], McGranahan *et al* [1972] and Rummel [1972]). Kamrany and Christakis (1970) also cite it as being necessary and one of seven sufficient conditions for the development of theoretically based social indicators.

The principles of parsimony and completeness together demand that any set of externally valid indicators of an education system should exhaust the definition of the space and do this in the most economical way. There should therefore be an isomorphic relationship between the final indicator set and the definition of the education system space.

Indicators measuring conceptually similar features of an education system should not be included together in any indicator set defined for monitoring or evaluating system development if that set is to have high external validity. This requirement immediately raises the difficulty of defining and selecting particular indicators (and their component variables) for use in a particular situation.

If similar indicators are available, how is one to be chosen in preference to another? This problem is known as the 'interchangeability of indicators' problem. It is discussed in the next chapter.

2.7 Résumé

The measurement of the characteristics of education systems is an undertaking requiring decisions to be made along many different dimensions. The complexity of these systems implies that any measurement operation must be very carefully planned and conducted. It is certainly an operation of far greater complexity and difficulty than the more frequent psychometric-type operations which generally are considered as forming the main focus of educational measurement. The implications stemming from the data collected from education systems have much greater ramifications in all spheres than do test results, however. Political and social decisions, not just educational decisions, are made on the basis of interpreting such data. It is vital that the data are as valid and reliable as is possible, that they are measured on appropriate scales and that they measure appropriate features. This chapter has outlined some of the considerations to be made in the quest for good education systems data.

Chapter 3

The Theoretical Definition and Development of Indicators

As was indicated in Section 1.6, one of the main approaches to the development of indicators of education systems involves defining them theoretically. This approach can be followed in one of two ways. One method involves identifying a set of variables in some *a priori* way and subsequently combining them to produce a single value for the indicator. The other method to construct an indicator by theoretical definition is to adopt an already existing indicator or formula which presently serves a particular purpose. There is not necessarily a strict theory-oriented purpose implied by either of these methods, although where this purpose can be incorporated as well, further implications for the use of any such indicators immediately derive from the definition. Targets set as part of a plan for developing particular aspects of education systems can, however, equally well lead to indicators being defined theoretically.

3.1 Implications of alternative strategies in developing an indicator from a concept

When developing an indicator from a definition of a concept to be measured, the three steps discussed in Section 1.6 must be followed. These are:

- ☐ select the component variables
- ☐ determine the method of combination
- ☐ define the weights to be applied to each variable.

Decisions to be taken at each of these steps have certain implications for the final form of the indicator and for the manner in which values calculated for the indicator can be interpreted. It is important, therefore, that the implications be very carefully considered and assessed. Unfortunately many indicators developed for particular purposes have not had proper cognizance paid to the implications stemming from each of these three steps. Consequently the interpretation of values calculated for some of the theoretically defined indicators is not always possible in ways implied by the concept as originally defined. Hence, to establish a basis for later discussion, some general principles to consider when making decisions at each of the three steps are now elucidated.

Selecting the component variables

The selection of variables for combination to form a theoretically defined indicator can be effected in a variety of ways. Normally, variables which might be included can be easily identified. Making a selection from just these variables, however, can lead to a bias in the underlying construct measured by the final form of the indicator. A more systematic approach than this is required.

To ensure that all relevant variables are considered when a selection is made to define an indicator, it is necessary firstly to develop a statement concerning the sorts of characteristics the indicator is to measure. This statement should be unambiguous and precise. It should be formed without necessarily considering, at this stage, what data may be available for use. In some instances, such a statement is derivable from a planning document or a theoretical proposition. In other situations an expert consensus must decide exactly what characteristics of an education system are to be covered by a defined indicator. In arriving at such a consensus, a two or three dimensional grid delineating the main factors involved in the characteristic can sometimes be a useful aid. If a definition indicates that too many variables might have to be combined to form a single indicator, a very complex concept is indicated as having to be measured. In such a situation, it may be preferable to reduce this complexity so as to avoid confounding different aspects of the concept. If confounding occurs, interpretation is made more difficult. Whatever approach is adopted, the final definition must ensure the defined indicator is directly and closely related to the total education system concept to be measured.

Once the various characteristics to be included have been identified, variables can be selected or defined to measure the specific features of each characteristic. Multiple measurements of a single characteristic are not permissible if parsimonious definition is to be achieved. Also, variables suspected to contain large errors of measurement should not be selected for combination. It might even be better to defer temporarily the development of an indicator than to proceed with the formation of a distorted composite.

When a choice still exists among a number of variables after considerations such as those outlined above, selection can sometimes be assisted by inter-correlating different variables to determine the extent of their empirical overlap. Diagram 3.1 demonstrates a possible situation which could arise when selecting variables to combine into a composite index. There, variables 1 and 2 are represented as being highly interrelated to the concept being measured. Such a determination would have to be ascertained by judgement rather than by empirical means when the indicator has not yet been formed. Both of these variables are also highly inter-related. As they are therefore probably measuring a similar characteristic, they should not be both included in the variable set. Variable 3 on the other hand has a low correlation with both variables 1 and 2 and a moderate correlation with the concept. It therefore contributes to the definition in a more independent way. The indicator of the concept in this case would be best formed by combining variable 1 or 2 with variable 3. In many situations, this empirical check on the behaviour of variables cannot be employed. Then only expert opinion, guided by the purpose for which the indicator is being developed, can be used.

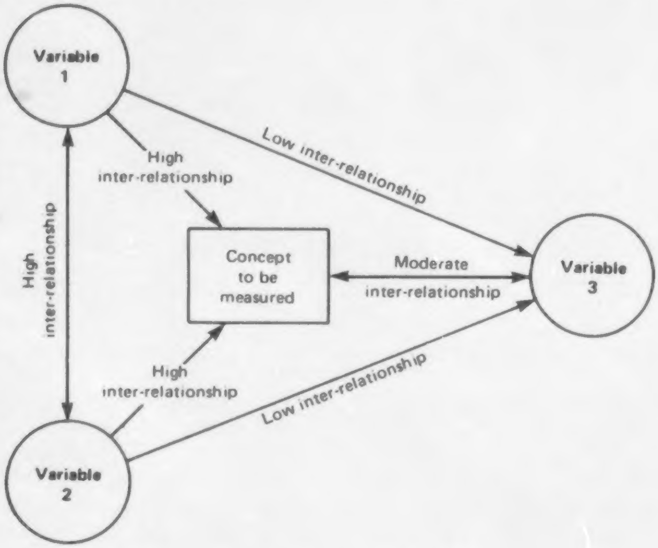

Variable 1
High inter-relationship
Low inter-relationship
High inter-relationship
Concept to be measured
Moderate inter-relationship
Variable 3
High inter-relationship
Low inter-relationship
Variable 2

Diagram 3.1: *Hypothetical inter-relationships among three variables and a concept to be measured*

Determining the method of combination

The vast majority of theoretically defined indicators are formed by simply adding together the component variables. Sometimes this is a most correct method of combination. At other times, it is not correct for addition does not always reflect the combination of variables as defined by the concept for the indicator. Multiplication may be a more appropriate technique of combination. In the early stages of indicator development, uncertainty about the proper method of combination often exists. As Coleman points out, 'if a precise theory existed, there would be no question about how these observations were to be combined to give the single measurement' (1964, 77). As development occurs therefore, refinement of techniques should also occur.

When addition[1] is the method used for combining variables, an implicit assumption is made that any one variable can substitute or compensate for another. For example, an indicator (I) might be formed from two component variables (V_1 and V_2) as follows:

$$I = V_1 + V_2$$

An average total value could be calculated for this indicator either by both component variables having average values or by one variable being considerably above the mean and the other being considerably below the mean. If both variables have similar distributions with means of 20 and standard deviations of 5, a value of 40 for I could be achieved by having $V_1 = V_2 = 20$ or by having $V_1 = 5$ and $V_2 = 35$. The latter case represents a wide diversity in level of the variables being combined, but the method of addition implies that compensation for the small value of V_1 by the large value for V_2 is acceptable. Values formed by addition are generally aimed at an interpretation in terms of differences – such as 'on indicator I, system A has a value which is z units higher than that for system B'.

Multiplying[1] the values for the component variables together to form an indicator implies a different concept from that of addition. An indicator formed in this way can be represented as:

$$I = V_1 \times V_2$$

Under this method of combination, compensation for a small value for one variable cannot occur as readily. The operation implies instead that a high indicator value can only be calculated when all component variables have high values. The value of one variable thus builds upon that of another. Using the same values as in the two situations described above for the operation of addition, the two indicator values formed by multiplication are 400 (= 20 x 20) and 175 (= 5 x 35). The low value for variable V_1 in the second instance cannot be compensated for by the high value for V_2. The indicator values therefore reflect that a higher level of the concept is present in the first situation than is present in the second. Indicator values formed by multiplication are generally formed when it is required to make statements like 'on indicator I, the value for system A is x per cent above that for system B'.

One implication of the difference between the two methods of combination is the way in which an average value can be formed. When the operation of addition is used, an arithmetic mean can define the indicator value.[2] That is:

---

1. Subtraction is regarded as the reverse operation of addition and division as the reverse operation of multiplication. Neither subtraction nor division need to be discussed separately therefore.
2. The median and the mode can also be used to estimate average values but as these can be subject to wide fluctuations especially in indicator formation,

$$I = \frac{V_1 + V_2 + V_3 + \ldots + V_n}{n}$$

or for weighted variables

$$I = \frac{w_1 V_1 + w_2 . V_2 + \ldots + w_n . V_n}{w_i} \qquad \textit{(Equation 3.1)}$$

where w_i is the weight for the ith variable V_i.

When multiplication is used, the geometric mean is appropriate. In its unweighted form, this can be expressed as

$$I = \sqrt[n]{V_1 . V_2 . V_3 . \ldots V_n}$$

and in its weighted form

$$I = \sqrt[n]{V_1^w 1 .\ V_2^w 2 .\ V_2^w 3 \ldots V_n^w n} \qquad \textit{(Equation 3.2)}$$

The computation of a value from this equation is facilitated by the use of logarithms.

The differences between the two methods of calculation are demonstrated in Table 3.1. There an indicator formed by two equally weighted variables is represented. The value of the arithmetic mean for the composite indicator in both year 1 and year 2 is 1.5. This allows discussion of the overall level of the indicator. Changes in the indicator value cannot be assessed by comparing the composite values in both years because differences in the values of the components are important and must be taken into account. As variable 1 increased 100 per cent and variable 2 decreased 50 per cent, an arithmetic mean calculated for the change shows a 25 per cent average increase in the indicator has occurred. Using the geometric mean on index numbers (so as to estimate the relative levels of both variables in year 2 – values of 200 and 50 respectively for the two variables) shows no change has occurred (= $\sqrt{200 \times 50}$). The increase by a factor of 2 in variable 1 is balanced by the decrease by a factor of 2 in variable 2. The geometric mean is thus more applicable in situations where ratios of change are to be considered. In indicator construction, each situation must be assessed and the appropriate method of combination applied to the component variables.

Variable	Value in year 1	 year 2	% change
1	1	2	+ 100
2	2	1	– 50

Table 3.1: *Use of arithmetic mean and geometric mean to measure change in indicator values*

---

they are not considered here. For a discussion on these statistics, in this situation, see Huff (1954) and Crowe (1965).

Defining the weights

The distinction between theoretically defining an indicator rather than allowing empirical definition is most marked when the problem of determining weights for the component variables arises. Partly because of the subjective nature of the decisions which have to be made, the definition of weights in this method of developing indicators can be a most contentious area. Its importance cannot be diminished because the selection of one set of weights rather than another can substantially change any results.

The most common approach used to select weights is to define all component variables as being equal in value and importance. Hence each variable is assigned unit weight. If this could always be done, no weighting problem would exist. The alternative to this approach is to use differential weights for the variables. Here planners and researchers nominate their own values for the weights – with or without justification. This is also a very commonly adopted procedure. Harbison and Myers (1964) do this with their composite index to be discussed in detail in the next section. Gulley and Newton demonstrate the approach very well in their development of an indicator to measure the distribution of leisure-time facilities:

> The authors felt that the youth population has the greater need for leisure time services. Accordingly, this factor was assigned a weight of 18 out of a total of 30 for the population group. Since youth population and population density are generally positively correlated, density was given a weight of 5. The weight of 7 was assigned to the aged population because the needs of the aged were considered to be greater than for other adults but less than those of the youth population. (1972, 190)

Expert judgement from sources other than the researcher can also determine weights. Sometimes this might be very simplistic and merely involve estimating percentages responding in particular categories of a questionnaire. In other situations, it can be very much more sophisticated. Dalkey and Rourke (1971) provide an example of such an approach. They use the Delphi technique to survey responses from a deliberately selected group of respondents. As subsequent information is progressively derived from the respondents and refined, a basis for the weights used to develop an indicator is formed.

A second approach to defining weights is to measure the effort required to establish the value of a particular variable. One such measure of effort is the time taken or allocated to produce a variable value. This approach assumes that if the time taken to do something is longer for one variable than for another, then that variable should be weighted more (or less) heavily. Hobson and Mann (1975, 451) define weights for a social indicator according to the time spent by a group of individuals on a particular activity compared to the total time that same group spends on activities of all types. Ferriss defines an index involving more weight being given to PhD degrees than to bachelor degrees because 'time is proportional to the student's effort' (1969, 264). If more time means poorer performance, a weight estimated by the inverse of the time taken can be used. Cost is another way of measuring effort to establish the value of a variable. Hence, cost could be used as a weighting factor. Consolazio (1967), for example, weighted degrees according to the estimated costs incurred by institutions providing the degree to derive an index of degrees awarded.

A third method of defining weights is to use agreed to standards. In education, few opportunities overtly present themselves to allow the identification of standards. Guidelines for minimum standards in some areas of educational provision can, however, be deduced from many sources. Policy documents, planning schemes as well as some research studies could all be used for such a purpose. In non-educational settings, standards are often adopted to define weights for indicators. Shults and Beauchamp (1972), for example, follow this approach in developing an air quality index. They nominate four different pollutants as being the variables to be combined to form their indicator. This indicator is to demonstrate that 'an intolerable situation exists when the average concentration of a pollutant attains or exceeds the standard value' (1972, 202). To achieve this, the weights assigned to each pollutant should 'reflect the importance of a given pollutant in a given region . . . [and] also take account of the normal fluctuations in pollutant levels' (1972, 202).

Shults and Beauchamp calculate the weights for their four pollutants from the following formula:

$$W_i = \frac{\sigma_i}{S_i - \mu_i}$$

where W_i = weighting factor
μ_i = the mean value for the level of the ith pollutant over a given period
σ_i = the standard deviation of the level of the ith pollutant over a given period
S_i = the standard or critical value as recommended for the ith pollutant.

The standard values adopted by Shults and Beauchamp for their four pollutants were those recommended by the US Environmental Protection Agency as being the secondary standard concentrations. Hence the above formula estimates for each region being considered, a weight which reflects the relative concentration of a pollutant relative to its standard or critical value. The higher the value calculated, the greater the concentration of the pollutant in that area and hence the indicator value reflects a more unhealthy environment.

Limited empirical methods can also be used on occasions to guide the decision on weights. Rank on a series of tasks is one such method. Frequency of occurrence is another. Angell (1951) demonstrates this latter consideration in constructing a crime index. He establishes separately the ratio of the rates for robbery and manslaughter to that for burglary and then takes the square root of each of these ratios. The resulting value is the value of the weight for each of the variables. This gives:

> the crimes in the three categories weights proportional to the square roots of their frequencies. One homicide equals, not one burglary, but 7.75 burglaries. This seemed a justifiable procedure since it appeared to be better than weighting one homicide as equal to one burglary or than weighting all homicides as equal to all burglaries. (1951, 60)

Such a procedure is particularly useful when, as in Angell's case, variables referring to quite different kinds of observations are being combined. Correlations (or their squares) between component variables and other factors known to be important can also form the basis of weights. Coefficients estimated for variables from a regression equation can also be used for weights in some situations (see Section 4.1).

The selection of one method of determining the weights for theoretically defined indicators rather than another depends on many considerations. The nature of the indicator itself must be taken into account; so too must the nature of the variables being combined. No definite rules can therefore be proposed to guide indicator development of this type.

3.2 Norm-referenced indicator formation

Most education system variables are measured on either ratio or interval scales. The first methods to be described below assume such scales. They are among the most frequently employed methods for theoretically deriving indicators. These methods are also the most straightforward and easiest to apply. For this reason, it is sometimes tempting to use them even when the underlying scales are nominal or ordinal. Sometimes the more powerful procedures can be acceptable in such situations. Lord (1953) for example, shows that proper interpretation is possible of mean and standard deviation values calculated from nominal categories (in his case, the number of each player in a football team). The adoption of the more powerful procedures must, however, be effected with full knowledge of the dangers involved. If these dangers are not recognized, interpretation of the values derived for the defined indicators will be confused and misleading at best. It must be recommended, therefore, that only methods appropriate to each type of scale are employed on those scales.

Non-distributive indicators from ratio or interval scales

The most elementary approach to index formation is that of simple addition. It is the approach most commonly thought of when indicator development is discussed.

The method and its pitfalls are demonstrated in Table 3.2. There the 1972 percentage of students who are female at each level is reported for six countries. An indicator can be formed by adding together the three percentage values for each country. The composite score is reported as 'index 1' in the tables. For the African countries reported, Egypt is shown to have the highest value for the female involvement indicator; Ecuador has the highest indicator value of the three South American countries listed. The indicator values show some differentiation among the African countries but minimal differentiation among the South Americans. The values reported can no longer be interpreted as percentages. Dividing the composite by 3 does give an average percentage although this value does not really mean anything because of the different quantities involved in the denominator of each percentage. In addition, the three levels are too diverse in what they imply and entail to permit any interpretation of a calculated average value. The value is thus useful only for ranking the countries.

Region	Country	% of students female			Composite index		
		1st-level	2nd-level	3rd-level	1	2	3
Africa	Egypt	38	33	28	99	307	505
	Rwanda	45	34	11	90	304	508
	Zambia	45	33	15	93	309	507
	Mean*	39.5	31.1	18.9			
	Standard deviation	7.3	10.6	15.5			
South America	Columbia	50	48	27	125	308	500
	Ecuador	48	48	30	126	304	496
	Guyana	49	49	26	124	305	501
	Mean*	48.1	49.0	36.6			
	Standard deviation	2.1	3.9	7.7			

* The means and standard deviations are calculated across all countries in each region and not just across the countries reported here.

Source: Unesco *Statistical Yearbook*, 1974 and 1975

Table 3.2: *Indicator formation by the method of simple addition – female involvement index*

The major problem preventing interpretation of the composite indicator formed by simple addition in this way is that changes in percentage points reflect different development efforts. Table 3.2 also reports the means and standard deviations for each of the three variables calculated separately for each region across almost all countries listed by Unesco in each region as having educational provision at the three levels. The standard deviation statistic gives an indication of how much variability there is among the nations. The values reported in Table 3.2

show that among the South American countries, a change of one percentage point at the third-level is 3.7 (= 7.7 ÷ 2.1) times more influential in forming a composite by simple addition than is a corresponding change at the first-level. The larger variability in both regions in the third-level distribution compared to the variability for either of the other two levels has the effect of magnifying differences among values at that level. These differences cancel out small yet significant differences at the other levels.[1]

To overcome this problem associated with simple addition, it is necessary to transform the observed values so that the variables forming a composite index can be weighted equally or in some pre-determined ratio. This can be achieved in various ways. If, for example, all three variables in the female involvement index are defined as contributing equally, the standard deviations for the three variables must be equated[2] and the adjusted values for each country added together.

The result of such a transformation is recorded in Table 3.2, as index 2. Interpreting this index leads to different conclusions about female involvement in education. Zambia is now shown to have the highest indicator value of the African countries and not Egypt. This has happened because although Zambia is seven percentage points ahead of Egypt on the first-level variable – a difference of 96 per cent of one standard deviation – it is 13 percentage points behind Egypt on the third-level variable – a difference equivalent to 84 per cent of a standard deviation. Equating the variability of the distributions ensures the greater relative difference at the first-level is now taken into account. For similar reasons, Columbia displaces Ecuador as having the highest index 2 value amongst the South American countries listed.

If the three variables in the index are not considered to be equally important, two alternatives are available to allow the index to be redefined. One alternative would be to accept the differences in variability with their implication that the indicator is influenced most by the observed values for the third-level. If this alternative is adopted, there is then no real reason to maintain all three variables in the definition. A representative index might be formed by selecting just one of the three percentage values. Leaving the index as it is implicitly has the same effect. The scope of the indicator is then diminished as too is

---

1. The fact that, for the African countries, the average values for the second- and third-levels are lower than the average for the first-level has no effect on the result. Constant values could be added to each of these mean percentages to make all three values the same. This would not, however, have any effect on their variability.
2. In the present example, this can be approximately achieved for the African countries by multiplying the first-level values by 4, the second-level values by 3, and the third-level values by 2. For the South American countries, the first-level values are multiplied by 3.7, the second-level values by 2, and the third-level values can be left as they are. Terwilliger and Anderson (1969) show that such an approach as this is fairly accurate for most practical purposes.

its usefulness in any situation requiring comment on overall female student involvement in education.

The second alternative to the 'equally weighted' assumption posits that female involvement at one level is more important than at the other levels. On a cross-national basis, this would be a difficult argument to sustain. On an intra-national basis, it would be much easier because of statements presumably made in documents such as the official development plans. An assumption could, for example, be made for the present data, that second-level education is three times more important than either of the other two levels. The application of the appropriate weights[1] to achieve this defined index produces the values for composite indicator 3 in Table 3.2. Rwanda is now shown to have the highest level of female involvement of the African countries on the defined indicator, while Guyana is now the South American country with the highest indicator value. All three countries in both regions can thus be demonstrated to have the highest rank depending on the way the female involvement indicator is defined and calculated.

This example has not been designed to demonstrate the commonly quoted saying 'there are lies, damned lies and statistics'. Nor is it intended to infer that the production of indicators should only be attempted by a statistician – that is a person 'who draws a mathematically precise line from an unwarranted assumption to a foregone conclusion' (Levinson, 1963, 218). Instead the example demonstrates that simply adding together the values for two or more variables produces a result which can be difficult to interpret. The implications of defining an indicator in a particular way must therefore be very carefully thought through.

In education, some studies define indices as being the composites of selected variables. The study by Harbison and Myers is perhaps the best known of these. Here two variables – the adjusted second-level and the unadjusted third-level enrolment ratios – are selected and combined into what the researchers term a 'composite index'. The third-level provision is considered to be more important 'to distinguish among countries' (1964, 31) and thus 'higher education should be weighted more heavily than second-level in such an index' (1964, 32). Harbison and Myers hence select weights of 1 and 5 to apply respectively to the second- and third-level enrolment ratios.

In the original data reported, the standard deviations for the two ratios are 22.6 (for second-level) and 5.0 (for third-level). What the weights achieve, therefore, is the approximate equation of the two levels and not the making of one level five times more important in forming the composite than the other. Hence each variable contributes about 50 per cent to the value a particular country receives on the final index.

---

[1] The weights for the first-, second- and third-levels respectively are now: for Africa = 4, 9 and 2; for South America = 3.7, 6 and 1.

Standardizing scores

The problems involved with simply adding variables together to form a composite indicator or attempting to select multipliers which will equate the standard deviations can be substantially diminished by adjusting the distribution of each variable so that it has a pre-determined mean and standard deviation. Selected weights can then be applied directly to these transformed or standardized variables.

An example demonstrates the formation of indicators in this way. For those countries listed in Table 3.2, the observed percentage of females at each level is adjusted by subtracting from it, the mean value for each distribution and then dividing this difference by the corresponding standard deviation. In particular, the equation for this transformation to produce a standard score (usually represented by the letter 'z') equivalent to an original (ie observed) score (say X) can be written as

$$z = \frac{X - \bar{X}}{S} \qquad \textit{(Equation 3.3)}$$

where $\bar{X}$ = mean of the original distribution
S = standard deviation of the original distribution.

For example, for Egypt, the percentage for the second-level is standardized as follows:

$$\text{Standard score} = z = \frac{33 - 31.1}{10.6} = 0.18$$

The standardized value[1] for each variable and country in Table 3.2 is reported in Table 3.3. These values are commonly referred to in the educational and psychological measurement literature as 'z-scores' (see, for example, Terwilliger [1971], Ebel [1972, 485] and Mehrens and Lehmann [1973, 150]). The z-score distribution has a mean of 0.00 and a standard deviation of 1.0. Adding the z-scores calculated from Table 3.2 together produces an index score as recorded for 'composite index 1' in Table 3.3. Weighting the value for the second-level as being three times as important as the values for each of the other two levels and adding these weighted z-scores produces 'composite index 2' in that table. The same interpretation can be made of indicators 1 and 2 in this table as was made of indicators 2 and 3 respectively in Table 3.2. This is because the only difference between the method used for Table 3.3 and the previous one is that the standardizing methods are more precise and accurate for Table 3.3.

---

[1] In no way does this formula for standard scores impose a normal curve on the distribution of the variable. The shape of the original distribution is completely preserved. Change of shape, when this is desired, requires a different score transformation.

Region	Country	Standardized values for % of students female			Composite index	
		1st-level	2nd-level	3rd-level	1	2
Africa	Egypt	−0.21	0.18	0.59	0.56	0.92
	Rwanda	0.75	0.27	−0.51	0.51	1.05
	Zambia	0.75	0.18	−0.25	0.68	1.04
South America	Columbia	0.90	−0.26	−1.25	−0.61	−1.13
	Ecuador	−0.05	−0.26	−0.86	−1.17	−1.69
	Guyana	0.43	0.00	−1.38	−0.95	−0.95

Table 3.3: *Indicator formation by the method of adding standardized variables – female involvement index*

The method of indicator formation described here can be summarized for a particular country by the equation:

$$I = w_1.z_1 + w_2.z_2 + w_3.z_3 + \ldots. + w_n.z_n \qquad \textit{(Equation 3.4)}$$

where I = indicator composed of n variables
w_i = weight for component variable i
z_i = standardized score of variable i.

The method of using standardized scores and weighting these scores before adding them to form the final value is the method implicitly underlying almost all of the techniques to be described in the next chapter which can be used to form indicators by the method of empirical definition. A very precise estimate of the value a system should be assigned for an indicator is produced by this method. That value, of course, must be interpreted in conjunction with the appropriate qualifications required by the errors of measurement involved with the original observation.

For a variety of reasons, it is generally preferable to divide the value calculated by Equation 3.4 by the number of component variables which have been included in the calculation (as is described in Equation 3.1). The average value so formed is then the indicator value. It offers a better value for comparing different indicators which may be based on different numbers of component variables. It also allows indicator values to be formed when the values for all component variables of an indicator are not available. In these circumstances, only those variables for which data are available (or considered to be reliable and valid) need be used. This added step of calculation is assumed to have been followed in the rest of the discussion in this section.

The range of values for a single variable calculated by the standard score defined by Equation 3.3 is usually from −2.5 or −3.0 to about +2.5 or +3.0. Values greater than these can certainly be calculated but normally such values identify countries which are classified as 'outliers' and whose values need special consideration in their estimation and interpretation (see, for example, Rummel [1970b, 280] or Johnstone [1977, 60]). Negative values are just as common as are positive values and indicator values should really be reported to two places of decimals.

To ease the phobia many people have about scores such as the z-scores reported in Table 3.3, as well as to simplify some reporting practices, the standardized observations as reported above for each variable can be transformed to what are known as 'T-scores'. These new scores conform to a distribution whose main statistics are imposed on them. Very often a mean value of 50 and a standard deviation of 10 are selected for a T-score distribution. To convert z-scores for a single variable to this particular T-score distribution, the following formula can be used:

$$T_i = 50 + 10.z$$

where T_i = T score for variable i
z_i = standard score for variable i.

The values for T_i for all i variables are then calculated and substituted into Equation 3.1 (the T_i values replacing the V_i values).

The general formula for estimating a T-score distribution can be expressed as:

$$T_i = \overline{X} + S.z_i \qquad \textit{(Equation 3.5)}$$

where $\overline{X}$ = value selected for the mean of the new distribution
S = value selected for the standard deviation of the new distribution.

Any values for the new mean and standard deviation can thus be chosen. The distribution of IQ scores, for example, is often based on a T-score distribution with a mean of 100 and a standard deviation of 15.

A less precise but more straightforward conversion from observed scores to a transformed score which generally approximates a T-score can be calculated from a score distribution as follows:

$$\text{Transformed score} = \frac{\text{observed score} - \text{lowest score}}{\text{range of distribution}} \times \frac{100}{1} \qquad \textit{(Equation 3.6)}$$

This formula obviates the need to calculate values for the mean and standard deviation. If the original distribution is not too skewed and does not contain extreme outliers (which will give an unrealistic value for the range) the values provided by Equation 3.6 are perfectly acceptable for indicator formation. Little and Mabey (1972) use such an equation to standardize the values of their 10 component variables before adding them together.

In the form defined here, standardized observations – either in the form of z-scores or T-scores – have not been used very often to develop indicators. Cutright (1963) is one who does use T-scores most properly.

He relates an index of political development (discussed in the next section) to four indicators measuring level of national development. One of these four is an education system indicator formed by separately converting two variables – the number of higher education students per 100,000 population and literacy rates – to T-score distributions and then combining them. A communications indicator is similarly formed by combining four separate variables. Although to form his education index Cutright combines input variables with education system-society congruence variables (as defined in Section 1.5), he does use the correct method to form composite indicators for his four independent indicators of development.

One difficulty with converting the values for variables to standardized observations is that the value finally calculated depends on the distribution of values across the observations (or education systems) included in the original sample. With a different or modified set of countries, different means and standard deviations will be calculated. Different z-scores will therefore result. This difficulty can be seen in Table 3.2. The statistics for the African countries are quite different to those for the South American countries. Combining both regions would have given very different values for the statistics, and hence different z-score values and different indicator values. Subsequent results might also have been different. This difficulty is reiterated in the next chapter because it plagues all of the empirically defined indicators. Generally, however, it is not a serious weakness here. Its main implication is that interpretation must be made on the basis of comparing a particular score to others in the sample – that is in strict norm-referenced fashion – rather than on the basis of absolute comparison, or in a criterion-referenced way.

Non-distributive indicators from ordinal or nominal scales

Few education system indicators can be formed theoretically from variables measured on nominal scales. The nature of such scales and the restricted measuring ability they have preclude all but very basic indicators from being formed by theoretical definition. Perhaps the most common procedure for developing theoretical indicators from nominal scales is to convert them into a type of ordinal scale containing only two categories – YES and NO. The number of characteristics contributing to a concept can then be totalled and this number used as the indicator value. For example, an index might be defined to measure community involvement in school administration. The variables contributing to this index would be questions identifying, for each school system being measured, whether certain aspects of community involvement are featured or not. The questions could be derived by consideration of the type of involvement perceived as being most important, from the findings of previous research studies conducted in the area or from some other source. A sample of such questions is presented in Table 3.4. Summing the number of YES responses for a given system provides a value to compare with values derived from other systems. This is a most simplistic approach to indicator formation, however, and hence nominal scales seldom provide a good basis for indicator formation.

Identify those characteristics from the following list which indicate the extent and nature of community involvement in this education system.

The community outside of the school:

☐ participates in activities to raise minor funds	☐ YES	☐ NO
☐ helps in day-to-day school administration	☐ YES	☐ NO
☐ is consulted on curriculum decisions	☐ YES	☐ NO
☐ participates in selection committees for hiring new staff	☐ YES	☐ NO

Table 3.4: *Sample of features considered to ascertain the extent of community involvement in an education system*

Ordinal scales provide very much more flexibility and consequently the quality of the indicators formed is superior to that for indicators based on nominal scales. A typical situation involving ordinal scales for the component variables of an indicator is reported in Table 3.5. There three questions are reported which are to be combined to form a single indicator called 'satisfaction with the education system'. This indicator is thus classified as an output indicator.

Rate your opinion of the extent to which the XYZ school system has helped its students achieve each of the following objectives.

Objective	Rating of the extent to which objective is achieved		
To become effective members of the work-force	☐ very well	☐ fairly well	☐ poorly
To accommodate changes in everyday life induced by modern technology	☐ very well	☐ fairly well	☐ poorly
To develop proper inter-personal relationships	☐ very well	☐ fairly well	☐ poorly

Table 3.5: *Sample component questions for an indicator measuring the satisfaction of individuals with the education system*

One way to form an indicator value from the responses to the three questions is to score each response as:

3 for a rating of very well
2 for a rating of fairly well
1 for a rating of poorly

and then add the scores for each question together to form a value for a particular education system. A maximum value for the indicator is then 9 while the minimum value is 3. (This could easily be adjusted to ensure a 6 to 0 range if it was deemed to be more convenient.) A seven point scale is thus formed. For many intermediate values, scores can have ambiguous components, however. A score of 6, for example, can be obtained:

☐ by all three components being ranked as FAIRLY WELL
☐ by all three ratings being used once each across the three questions.

This equates, therefore, a FAIRLY WELL-FAIRLY WELL-FAIRLY WELL rating with a VERY WELL-FAIRLY WELL-POORLY or a 'very mixed attitude' rating.

Accepting such an ambiguous method of arriving at the same value might be quite acceptable in many situations. In other situations, however, an additive model, as is indicated earlier, might not be the appropriate method of combination. A multiplicative model may be more preferable. Multiplying the values on each of the three questions together produces ten values between 1 (= 1 x 1 x 1) and 27 (= 3 x 3 x 3) inclusive.[1] For these values, no such contradiction in combination values is tolerated. For example, three responses of FAIRLY WELL would score 8 (= 2 x 2 x 2) while the mixed attitude noted above would score 6 (= 3 x 2 x 1). Hence a FAIRLY WELL-POORLY rating is not averaged out as being a FAIRLY WELL rating but as something a little lower. The indicator value thus reflects a lower satisfaction with the education system. If the results of the multiplicative operation are considered as producing artificial differences along the scale, the ten possible numbers calculated can easily be ranked from 1 (= 1) to 10 (= 27).

Variations on these ways to analyse Likert scales are available and frequently used. (Some are employed for example, by Culyer *et al*, 1972.) These are the most common scales particularly for measuring the input preferences and output satisfaction of a society for education. In some situations, multiple observations are collected particularly when open-ended questions are used in conjunction with the closed scales. Difficulties immediately arise in processing results but these can be overcome by methods such as outlined by Angell (1951, 62).

---

[1] The set of possible values obtained by using the multiplicative model is 1, 2, 3, 4, 6, 8, 9, 12, 18 and 27.

The weighting problem

A problem of weighting does arise on occasions when one question is considered to be more important or central in the definition of an indicator than another. In cases such as these, the concepts of variability developed in the previous section in relation to ratio and interval scales again apply. Thus, multiplying scale values by a constant value does not necessarily ensure that a particular desired weighting is achieved. Sometimes an assumption can be made that the ordinal scale approximates an interval scale and that standard scores for each question can be calculated. Scores can then be weighted and added if that form of combination is the one considered to be appropriate.

One study assuming the weighting problem can be overcome by definition is that by Cutright (1963). He develops an indicator of political development by combining ratings of the characteristics of two different branches of government within a country. Cutright awards scores of 2, 1 or 0 to a country for one variable depending on the diversity of political parties represented in its government. Scores of 1, ½ or 0 are awarded to the country depending on the way in which its executive branch of government is formed. Adding scores on these two variables gives Cutright his index of political development. If each of the three categories in each variable had the same percentage of countries assigned to it, the first variable would then be twice the weight of the second variable in the composite index.[1] As it stands in Cutright's paper, however, the real nature of the political development index is unclear. No information on distributions is presented. Either variable could dominate the entire index – thus nullifying completely the purpose for establishing a difference in scoring methods. Without being able to ascertain the actual weightings, therefore, the index of political development based on component ordinal variables is difficult to interpret.

Distributive indicators

All of the methods for developing indicators discussed so far facilitate comment on the overall or average level of development within a system. They do not allow comment on the diversity of that characteristic throughout the system. The distinction implied here is that discussed in Section 1.3 where the concept of non-distributive indicators is contrasted with that for distributive indicators.

Data of a global type cannot have distributive indicators calculated from them. These data comment on characteristics of the education system viewed as a whole. Aggregate data measured on either an interval or ratio scale can, however, be reported in distributive and non-distributive form. For example, the variable 'percentage of qualified teachers at the first-level' is a non-distributive variable when reported for a state or national education system as a totality. If, however, the value for each state or province in a country is estimated and a standard deviation or some other measure of dispersion is calculated, a measure of the variability of the percentage of qualified teachers at the first-level within the state education systems of a country is obtained. This statistic can then be compared to the corresponding estimates for other years or from other countries and the relative extent of the uniformity or diversity ascertained. Such a comparison was demonstrated in Table 1.3 to show differences among the geographical regions of Indonesia.

---

[1] The standard deviation of the 'diversity of political parties' variable would be 1.0 while the standard deviation for the 'method of executive branch formation' is 0.5.

A measure of dispersion such as the commonly used standard deviation is perfectly acceptable when all values contributing to the calculation are considered to be equally important. This cannot always be the case, however. Continuing the above example, it may be that some states with small enrolments have special restrictions on them. It may therefore be wished to make these states less influential than, for example, the bigger states. In many situations, units based on education systems allow only an assumption of equal units and/or equal weightings to be made in the calculation of a standard deviation statistic – despite the fact that such equality might not be meaningful on educational grounds. When inequality of contribution is meaningful, the standard deviation statistic – or any of its derivative measures – cannot be properly applied.

Coefficient of variation

A further difficulty in using the standard deviation statistic is that changes in the average level of a characteristic can make comparisons based only on the variability of that characteristic misleading in the practical situation. For example, a standard deviation of $2000 calculated for a group of individuals having an average income of $4000 is very much more significant in the inequality it reflects than a standard deviation of $3000 calculated for a group of individuals who have an average income of $100,000. Hence, because the mean and the standard deviation are statistically independent measures,[1] relative inequality or variability cannot be assessed by using either measures independently. They must therefore be combined. This is effected by the 'coefficient of variation' which is defined as:

$$CV = \frac{S}{\overline{X}} \qquad \text{(Equation 3.7)}$$

where S = standard deviation of the distribution of a quantity
$\overline{X}$ = mean of the distribution of that same quantity.

---

[1] They are independent because the formula for the standard deviation involves subtracting the mean from all observed values in a distribution.

High values for this coefficient indicate high variability compared to the average value (ie a situation of inequality) whereas low values indicate comparatively low variability compared to the average value. Comparing the coefficient of variation values for education systems on the same variable or indicator thus allows an estimate to be made of the equality of that variable or indicator throughout the components of each system relative to the average levels which pertain. Such an approach reflects a particularly realistic measurement operation in the area of planning because planners are constantly seeking to change the levels of particular characteristics in their education systems.

Table 3.6 demonstrates the use of the coefficient of variation. There the percentage of six year olds in school is averaged for each of the four geographical regions in Thailand for 1960 and 1970. Standard deviations, calculated from the provinces within each region, are also reported together with the values for the coefficients of variation. If an objective relating to regional dispersion in the participation rates of six year olds is to be evaluated with these data, neither the mean nor the standard deviation can really be used.

Region	1960			1970		
	Mean	SD	CV	Mean	SD	CV
North	2.07	1.81	0.88	12.85	3.54	0.28
South	1.99	0.84	0.42	13.29	4.31	0.32
North-East	1.80	1.00	0.55	11.79	2.55	0.22
Central	3.89	3.45	0.89	21.38	9.20	0.43

Source: Calculated from Sudapraesert and Tunsiri (1977, 95-6)

Table 3.6: *Basic statistics for the percentages of six year old children enrolled in the four regions in Thailand – 1960 and 1970*

Comparing the mean values shows that for all regions, the average levels of participation rate have increased substantially over the 1960-70 period, but the variability among the provinces within each region has also increased substantially. Hence Thai planners appear to have achieved higher participation rates at the expense of increasing the disparity of the rates among the provinces within regions. An evaluation of progress must resolve this apparent conflict. The coefficients of variation show that for 1960, the South region had the least relative inequality among the provinces within it while the North and Central regions (the latter includes Bangkok) had the greatest relative inequality. By 1970, the inequality existing within the North and North-East had apparently decreased the most – by 68 per cent and 60 per cent in the two regions respectively. Inequality had decreased by 52 per cent in the Central region and by 24 per cent in the South. The North-East is now the region showing least inequality among its provinces, while the

Central region shows the most. Hence, although overall levels of participation are higher in the Central region, conditions in the North-East are very much better in the sense that they are more homogeneous — which is the objective assumed for this demonstration.

The coefficient of variation can be calculated on variables measured or assumed to be measured on at least an interval scale. Descriptive statistics, especially measures of equality, do not have meaning when calculated on measurement scales which only rank or classify entities (ie ordinal or nominal scales). The coefficient is also only really useful for comparing inequality on the same variable either across groups or across time. Hence to conclude that inequality of six year old dispersion rates is different from the inequality of the availability of funds within regions for capital works on the basis of their respective values for the coefficient of variation is invalid.

Williamson (1965) notes the comment made above concerning the bias introduced into the standard deviation calculation — and therefore into the coefficient of variation — when it is assumed that an equal contribution to the calculation is made by all states, provinces (or whatever the units are on which the calculation is based). He posits that in many situations, such an assumption is invalid because some units should be regarded as being more important or influential than others. He proposes an alternative formula for the coefficient of variation which incorporates a weighting component into the numerator. His formula is:

$$CV_w = \frac{\sqrt{\Sigma(Y_i - \overline{Y})^2 . \frac{f_i}{N}}}{\overline{Y}} \qquad \textit{(Equation 3.8)}$$

where Y_i = quantity being measured for the ith region
$\overline{Y}$ = average value of Y across all regions
f_i = population of the ith region
N = total population.

Williamson uses this formula to measure the inequality of *per capita* income across a number of countries and considers it to be a useful alternative statistic to facilitate cross-national comparisons.

Few studies have appreciated the potential of the coefficient of variation. Jencks *et al* (1972) is one study that does. It uses the coefficient to compare the extent of the inequality of different sub-groups of American society on each of a number of variables. Some of the variables used in the study are number of years of schooling, income and test scores. The coefficient of variation is computed for these variables so as to compare the relative inequality between males and females, between blacks and whites and for each sub-group across time.

In other circumstances, conventional measures of dispersion do not apply conceptually. For example, enrolments in each grade might be calculated as a percentage of the total enrolment at that level. This variable can be viewed as a measure of the extent to which repetition and drop-out rates affect enrolment patterns. In a particular country, five grades of formal education might be provided at the first-level. If enrolments are evenly spread throughout that system and the differences in the number of children in the population at each of the relevant ages are not too large, about 20 per cent of the total first-level enrolment could be expected to be enrolled in each grade. If however 35 per cent of the total enrolments are enrolled in the first grade with a further 30 per cent in the second grade, considerable inequality in the distribution of enrolments is revealed. This in turn might be explained in terms of the effects of drop-out and repetition on educational provision. Many students might be retained in the early grades with relatively few progressing on to the later grades. To measure the inequality described here, statistics such as the standard deviation are not appropriate on conceptual grounds.

Lorenz curves

Many measures of inequality can be calculated. Jahn *et al* (1947) and Alker and Russett (1966) discuss some of these. One way to represent inequality in a diagrammatic way is to graph a Lorenz curve. It is formed by plotting the cumulative percentage of a characteristic against the cumulative percentage calculated or estimated for an appropriate criterion measure. In a situation of perfect equality, the corresponding cumulative percentages for the characteristic and the criterion should be identical. When graphed this produces a straight line referred to as the 'line of equality'. When an uneven distribution is present, the cumulative percentage of the characteristic is greater than the corresponding cumulative percentage of the criterion. Plotting the pairs of values for the characteristic and the criterion thus produces an arc in the space below the line of equality. This arc is referred to as the Lorenz curve.

An example demonstrates the formation of a Lorenz curve and its use to depict a norm-referenced comparison. The characteristic to be measured is the distribution of enrolments across grades at the first-level. The criterion is the grade – or, in particular, the percentage of total enrolment each grade would have if total enrolment was evenly distributed across all grades. If a country with six grades of first-level education had a perfectly equal distribution of enrolments across the six grades, 17 per cent of first-level students would be enrolled in grade 1, 33 per cent would be enrolled in grades 1 and 2, 50 per cent would be enrolled in grades 1 to 3 and so on.

Diagram 3.2 shows the plot of the cumulative percentage by grade of actual first-level enrolments against grades for each of El Salvador and Nicaragua. Table 3.7 reports the data for 1965 on which the graphs are based.[1] Also shown in the diagram is the line of equality which would

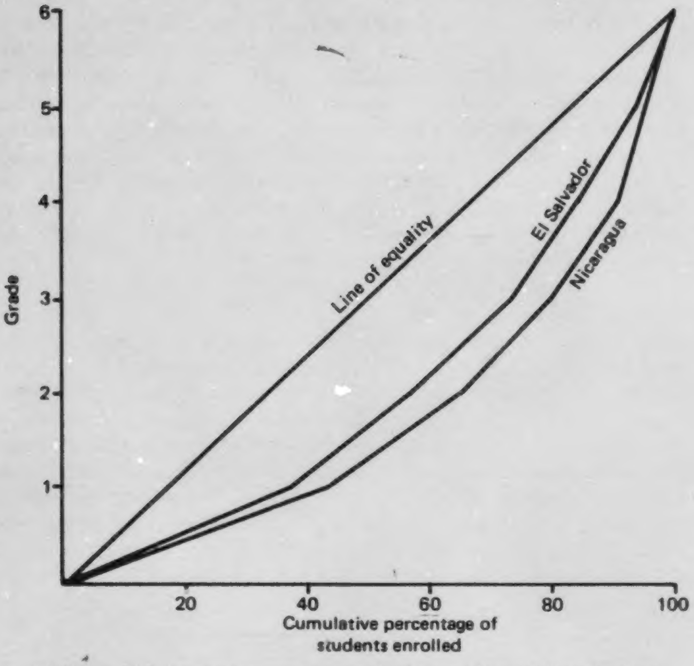


Diagram 3.2: *Lorenz curves measuring inequality of distribution of enrolments across grades at the first-level – El Salvador and Nicaragua, 1965*

Grade	El Salvador	Nicaragua
1	37	45
2	20	20
3	15	14
4	12	10
5	9	6
6	7	5

Source: Unesco *Statistical Yearbook* 1974, p192

Table 3.7: *Percentage distribution of first-level students by grade – El Salvador and Nicaragua, 1965*

---

1. The interpretation really implied by Diagram 3.2 is that, for El Salvador, 37 per cent of the first-level enrolments receive 17 per cent of the education offered at that level (ie first grade), 57 per cent of the total enrolment receives 33 per cent of the total education offered (ie first and second grades) and so on.

result if the distribution of the characteristic equalled the distribution of the criterion at each grade level. The two Lorenz curves appear below this line, thus indicating that in 1965 neither El Salvador nor Nicaragua had an equal distribution of enrolments across the grades provided at the first-level. As the curve for El Salvador is the higher of the two plotted, it may be concluded that there was greater equality in 1965 in the distribution of enrolments across grades in El Salvador than prevailed in Nicaragua. If this is due to lower repetition and drop-out rates in El Salvador, higher quality of instruction at the first-level in that country might be able to be inferred.

The Gini coefficient

When Lorenz curves are compared in this way, it is an easy matter to determine which education system has the most equal distribution of the characteristic being measured. Indeed, Bowman notes: 'The Lorenz curve . . . is a fairly sensitive indicator of inequality insofar as disparities may legitimately be weighted in terms of arithmetic differences' (1951, 99). When a large number of situations are being compared simultaneously, however, a more analytical method is required. The best of the available methods for synthesizing information from a Lorenz curve is the Gini coefficient. This ratios the area between the curve and the line of equality to the total area under the line of equality. Expressed mathematically:

$$G = \sum_{i=1}^{n} (p_{i-1}q_i - p_i q_{i-1}) \qquad \textit{(Equation 3.9)}$$

where p_i = cumulative proportion of the characteristic whose equality is being investigated

q_i = cumulative proportion of the variable which is acting as the criterion for the measurement.

The greater the equality of a particular characteristic, the closer the Lorenz curve is to the line of equality and the closer the value of the Gini coefficient is to 0.0. The greater the inequality, the further away the Lorenz curve is from the line of equality and the value of the Gini coefficient approaches 1.0.

The values[1] for the two Lorenz curves in Diagram 3.2 are:

for El Salvador, G = 0.31
for Nicaragua, G = 0.41

These values lead to the same conclusion as was reached above — that there was a greater degree of equality in El Salvador in the distribution of enrolments across grades at the first-level than there was in Nicaragua in 1965.

---

[1] The calculation for the Gini coefficient for El Salvador is reported in full in Appendix A together with technical details for computing Gini coefficients. An alternative computing formula is also provided.

In interpreting a Gini coefficient, particularly when making cross-national comparisons, two properties of the coefficient must be kept in mind. One is that the values calculated in no way comment on the level of the characteristic being measured. They comment only on the distribution of the characteristic throughout the region being considered. Hence two education systems might produce identical values for a Gini coefficient measuring, for example, the equality of female involvement at the first-level. All this implies is that the cumulative percentage of the total first-level female enrolment estimated over regions is approximately the same overall in both systems compared to the cumulative percentage contributed by each grade to the total enrolment. It does not imply that just as many girls are enrolled in each system.

A second property of the Gini coefficient is that it does not indicate where inequality is located. It merely reveals that inequality is present. Again, in the above case for Nicaragua and El Salvador, for example, the inequality across grades could have occurred in either country at any or all grade levels. Hence, if Nicaragua had a system involving public examinations at the grade 4 level and this caused the unequal distribution (because promotion from grades 1, 2 and 3 were relatively automatic) a similar value for the Gini coefficient could have been calculated. In most situations, not knowing where points of inequality are located is not a problem. The original data can always be consulted if and when such identification is required. In addition, the appropriate axis of the Lorenz curve should be labelled to indicate grades, provinces, etc contributing to that part of the curve. This helps to point to locations where inequalities might be greatest. Diagram 8.1 is an example of such a representation.

Many studies outside of education use Lorenz curves and/or calculate Gini coefficients. As the concept was developed in economics, it is a commonly used technique for measuring inequality there. Russett *et al* (1964) in their *World Handbook of Political and Social Indicators* report two cross-national tabulations of income distribution data based on Gini coefficients. Taylor and Hudson (1972) follow this example in the second edition of the *Handbook*. Jahn *et al* (1947), Bowman (1951) and Alker and Russett (1966) all investigate the effectiveness of the Gini coefficient *vis-à-vis* other related measures of inequality including some also stemming from Lorenz curves. All of these studies consistently find that the Gini coefficient is at least as good if not better than the other available indices of inequality. In addition they note that the Gini does not have some of the methodological disadvantages inherent in many of the other indices.

Unfortunately neither the Lorenz curve nor the Gini coefficient have really been used in educational planning in any recognized way. Alker (1965, 51-3) uses a Gini coefficient by way of example to measure racial imbalance in four junior high schools in the US and to ascertain the effect a school board proposal would have in diminishing that imbalance. Kottis and Kottis (1973) draw a Lorenz curve to show the extent of the disparity in educational achievement levels across various communities in Norway. They do not however elaborate on their diagram nor calculate any coefficients. Ferriss (1969, 216) compares across varying years, the educational attainment of 20-24 year old non-white males in the US by estimating the area under various curves he draws. His is an exercise with a similar conceptualization to that of drawing a Lorenz curve although the situation is defined differently.

The Gini coefficient can be used to measure the degree of equality of many characteristics of education systems. Sometimes such measures indicate a situation needing to be improved. In other instances, the measure merely comments on a proper situation which planners might hope continues to exist. Inequality of a characteristic does not therefore always signify a bad situation. The potential of a measure such as the Gini coefficient is considerable and it has yet to be exploited. Some situations in which the Gini coefficient may be calculated include the estimation of:

- ☐ the degree of female equality in enrolment across regions
- ☐ the distribution of over-age students across grades
- ☐ the distribution of enrolments at each level across geographic regions
- ☐ the distribution of teachers according to teaching experience across geographic areas
- ☐ the extent to which teachers at various levels of qualification are distributed across geographic regions of a system
- ☐ the extent of racial or religious balance in the education system.

The potential of some of these suggestions is explored further in Chapter 5.

The measurement of concentration

Many instances occur, particularly in monitoring education systems performance or comparing many systems simultaneously, when dispersion of a different kind to that measured by the Gini coefficient should be measured. These arise when it is required to measure the distribution of a variable across a number of categories which form either a nominal or an ordinal scale. To what degree, for example, is an education system provided with funds from a central source? What dependency is there on a single authority for providing education within a country? What is the degree of racial or religious mix in a particular area? In situations such as these, neither a Gini coefficient nor one of the other traditional measures of dispersion is suitable. No assumption can be made here about equal intervals. Nor is the problem one of examining the distribution of a variable *vis-à-vis* the distribution of a criterion. What can be calculated, however, is an indicator measuring the degree of concentration within a number of defined categories. Many such indicators have been proposed in the literature of the various fields in the social sciences. Taylor and Hudson (1972) and Hall and Teideman (1967) review some of these measures of concentration.

The frequently used measure of concentration is that known as the Herfindahl-Hirschman index. It is very simply defined as:

$$C = \sum_{i=1}^{k} p_i^2 \qquad \textit{(Equation 3.10)}$$

where p_i = proportion of total observations in the ith category
k = total number of categories.

If one category contains all observations, the C measure of concentration takes the value of 1.0. If all k categories in a defined scale have an equal proportion of the total amount – so no point of concentration exists – the C measure takes the value 1/k. When there are many categories, C is approximately 0.0 therefore.

An example demonstrates one way in which the C measure of concentration might be used. Table 3.8 reports for both 1965 and 1971 for four Asian countries, the percentage of third-level students in each of the nine fields of study classified by Unesco. Also reported are the indices of concentration as calculated by Equation 3.10. The 1965 C values show that Cyprus and Laos were both moderately concentrated in their third-level enrolments, especially in comparison to Japan and Saudi Arabia. By 1971, third-level enrolments in Cyprus had become very much less concentrated because of the introduction of courses in the social sciences and engineering. This change is reflected in the significant decrease in the C index of concentration to a level about equal to that calculated for Japan. It may be concluded therefore that the concentration of third-level enrolments in the two countries in 1971 was at about the same level despite the fact that Japan offered three more faculties than did Cyprus. The measurement of the degree of concentration is dominated in the case of Japan by the very high proportion of students still attracted to the social science faculties. Laos became slightly more diverse over the 1965-71 period because of the relative movement of students away from law towards education and medical science faculties. Saudi Arabia diversified considerably especially through a de-emphasis on courses in the humanities and law. Enrolments were therefore more evenly spread across faculties in 1971 than they had been in 1965. The values for the C index of concentration reflect all of these changes.

The C index in no way attempts to reveal which category is most important nor how many categories have been used. It is merely addressed to estimating the evenness of the distribution of a particular quantity – third-level enrolments in the above example – across a

Field of study	Cyprus		Japan		Laos		Saudi Arabia	
	1965	1971	1965	1971	1965	1971	1965	1971
Humanities	33.33	36.54	14.35	13.59	–	–	30.24	20.49
Education	–	–	8.02	9.97	9.32	14.25	4.30	14.30
Fine arts	–	–	1.91	2.58	–	–	–	–
Law	–	–	–	–	47.83	37.50	36.94	25.89
Social sciences	–	5.56	45.10	43.82	–	–	14.79	14.79
Natural sciences	–	–	3.64	2.61	–	–	4.34	7.46
Engineering	–	24.57	18.32	20.68	–	–	7.04	12.12
Medical sciences	57.64	30.25	4.91	3.56	42.86	48.25	0.78	2.83
Agriculture	9.03	3.09	3.75	3.21	–	–	1.60	2.11
Index of concentration	0.45	0.29	0.27	0.27	0.42	0.39	0.26	0.17

Source: Calculated from Unesco *Statistical Yearbook* 1974, Table 5.2

Table 3.8: *Measures of concentration and percentage enrolments by faculty – selected Asian countries – 1965 and 1971*

number of defined categories. To determine which category contains the highest proportion of that quantity, it is necessary to consult the actual data from which the index is constructed. Similarly if change is to be measured, the C measure indicates whether or not change has occurred and, if so, its direction and its extent. It does not indicate in which category or categories the change occurred. The index is thus a very proper global measure summarizing a characteristic of an education system in the way required of an indicator in Chapter 1.

One feature of the index of concentration is demonstrated by the data reported in Table 3.8. It is that if a characteristic is classified into a fewer number of categories in one system than it is in another, this dissimilarity *per se* can change the index value. In Cyprus, for example, the inclusion of two more fields of study by 1971 helped to reduce the calculated C value. By 1971, therefore, higher education in Cyprus was more diverse. Simply adding more categories does not necessarily lead to a reduction in the C value, however. When additional categories are added and they account for only a small proportion of the total distribution, the concentration of the characteristic being measured will not be significantly reduced. The comparison between the 1971 values for Japan and Cyprus reflects this.

Two further points need to be made concerning the index of concentration. One is that if many categories are defined for the distribution, only those 10–12 (at the most) containing the highest proportions of the characteristic being measured need to be used in the calculation[1] of

---

[1] Using only the highest percentages in this way simulates another commonly used measure on concentration. This measures the proportion of a population accounted for by the largest 3, 5 or 10 categories.

C. These should provide a fairly close approximation to the actual value for C and one which would be sufficient for all practical purposes — especially considering the errors of measurement inherent in measurement scales. Arriaga (1970) demonstrates the errors introduced by such an approximation are quite small. He also shows their theoretical magnitude by invoking the Riemann-Zeta function.

The second point to note is that interpreting the value for C and ensuring it is comparable across distributions for different years or for different education systems requires a fairly strict definition of the categories used. In the above example, students should not be categorized into humanities faculties by one country and into education faculties by another country. Similarly each category along the nominal scale — for this is the scale on which most measurements will be taken — must be defined clearly and must be mutually exclusive. Measurements involving multiple observations cannot be accommodated by the C index.

Hall and Tiedeman (1967) criticize the C index as defined by Equation 3.10 on the grounds that it weights each category by the share it has of the total quantity being distributed. If there are few categories, therefore, (as is the case with Laos in the above example) it can be argued that entry is harder and C is higher, and that this is not desirable. They therefore propose an alternative index based on weighting each proportion by its rank across the categories. The values obtained from their formula are, for all practical purposes, almost identical to the results provided by the computationally simpler Equation 3.10. The extra effort of ranking does not therefore seem to be warranted.

A very similar formula to Equation 3.10 has been proposed in other disciplines but with a different connotation. It is:

$$F = 1 - \sum_{i=1}^{k} \left[\frac{n_i}{N}\right] \left[\frac{n_i - 1}{N - 1}\right] \qquad \textit{(Equation 3.11)}$$

where n_i = number of observations in the ith category
N = total number of observations in the distribution.

Where n and N are large, the value for F is approximately that for (1 − C). This index has been called an 'index of fractionalization'. It measures the probability that two members, selected at random from the population being considered, do not have the same characteristic. From this point of view, the index of concentration C measures the probability that two members of a population selected at random have the same characteristic. They are, for example, enrolled in the same field of study. The C values in Table 3.8 show that it was much more likely to choose randomly in 1965, two Cypriot students studying in the same area, than it was in 1971.

The C measure of concentration has been used in a number of different areas. Taylor and Hudson (1972, 366-71) use it to measure the concentration of reports. Ivanovic cites a modification made by UNCTAD to the C measure to examine 'the concentration of births in relation to the age of the mother' (1974, 21). Arriaga (1970) uses the C index to measure the concentration of urbanization in a country. He also proposes an alternative measure equal to N times the C value (where N is the total number of people or observations being considered). He claims such an indicator is a better measure because it incorporates a size component into its value. This it certainly does, but by so doing, it contradicts the third recommendation for indicators established in Section 1.3 – that indicators must be independent of the potential of a country to provide a particular quantity. Arriaga's alternative form cannot therefore be considered further here.

The F index has also been used in a variety of ways. Rae and Taylor (1970) use it to measure what they term 'political cleavage' – that is the extent to which seats in a legislature are distributed among political parties. Greenberg (1956) uses it to measure linguistic heterogeneity in a population. Taylor and Hudson (1972) cite other instances of its use.

In relation to education systems, the index of concentration has not been used to any extent. It does have a great deal to offer, however, in helping to ascertain whether a plan proposal is likely to have a significant effect on the nature of an education system. Values for either the C or the F coefficient could be calculated as appropriate for the present period as well as for the situation desired by the end of a plan. Comparing the two values would reveal the extent to which plan implementation would change the nature of educational provision in that part of the education system being considered. Many possibilities exist for the calculation of indicators of concentration for planning, monitoring or research purposes. Some of these possibilities are:

☐ to determine change in concentration of educational funds flowing from central, state and local governments as well as from private and other sources
☐ to index the concentration of responsibility for providing education among the various authorities (government, church, other private, etc)
☐ to identify the extent to which teachers at various levels of qualification are distributed within geographical regions of a system
☐ to measure dispersion of ages of pupils within particular grades
☐ to determine the similarities in distribution of enrolments across grades between different geographical areas.

Some of the suggestions in the above list parallel those made in relation to the Gini coefficient. Comparing both sets of suggestions reveals the essential dissimilarity between the characteristics measured by the two types of indicator. The Gini coefficient could be used to measure the extent to which qualified teachers are distributed equally across geographical regions. Its Lorenz curve plots the cumulative percentage of all qualified teachers employed within each area against the cumulative percentage of all teachers employed within the corresponding area. For equality, therefore, each area would have the same percentage of total staff who are qualified teachers as its percentage of the total teaching staff. It would be inappropriate to use a measure of concentration for this purpose because it is known in advance that as the number of teachers in each area differs, the percentage of total qualified staff contributed to the total number by each area must also differ. What can be measured by an index of concentration, however, is the extent to which teachers' qualifications within a region are concentrated at one level or whether instead teachers within an area have a diversity of qualifications. In such a situation a high value for C would be ideally required – all teachers being properly qualified.

3.3 Self-referenced indicator formation

Indicators developed to allow self-referenced measurement of an education system focus either on a single system or on a very limited selection of a few education systems. Generally, therefore, they are different from indicators facilitating norm-referenced measurement. One reason precluding the use of some of the methods reviewed in the previous section is that no group statistics can be reliably calculated to form a basis for combining different variables into a composite indicator or to facilitate interpretation.[1] Statistics such as the mean and the standard deviation, for example, cannot be used. Either these cannot be calculated or they would have to be calculated from such a select and often biassed group of education systems as to be not sufficiently valid for general measurement purposes. Additionally, as planners work towards change, measures of central tendency and dispersion calculated on an education system deny planned system development. Further, the basic descriptive statistics differ considerably in value when calculated over different time periods.

The general problem raised here for developing indicators for self-referenced purposes has two distinct facets: against what basis changes can be measured; and how variables measured on different units can be combined. Once these issues are solved and a proper basis is established to ensure variables can be properly and meaningfully combined, methods of weighting and combination as described for other forms of indicator can be used.

---

[1] This in no way implies that indicators developed by norm-referenced methods cannot be used for self-referenced interpretation. They certainly can be. The extra effort usually required to enable this to occur, however, is only worthwhile in certain research contexts.

116

Non-distributive indicators

Index numbers can be used to overcome both aspects of the problem of developing indicators for self-referenced purposes. They are founded on the concepts of establishing a base point for each variable and the relating of all measurements back to that point in order to measure the amount of change from that base. In self-referenced measurement, the base frequently adopted is a particular year. Sometimes one year might be selected because it marks the beginning or the end point of a planning period. At other times, a base year might be a census year. Indeed any convenient year can be selected. The values for all variables — and hence for all indicators — in the base year are then set at a base level. A value of 100 is conventionally chosen. The values for variables in years either earlier or later than the base year can then be computed by relating them to the base value. Index numbers with values less than 100 reveal a decrease from the base year value, while values greater than 100 reveal an increase over the base year value. The difference between the value of an index number and its base year value is a measure of the percentage change which has occurred over the period from the base year.

An example demonstrates the main concept of using index numbers. Table 3.9 reports Unesco data for the component variables of an indicator measuring the human resources involved in the provision of education (as discussed in Chapter 5). The country in the example is Bulgaria and the data refer to the 1960-72 period. Also reported are the index numbers for each variable formed by ratioing the value for each year to the 1960 value for that variable. For example, the 1970 index number of 131 for the second-level female enrolment ratio results from the operation (59 ÷ 45) x 100. Also reported are two composite indicators — each formed by a different method. Indicator 1 for each year is formed by calculating the arithmetic mean averaged over the six index numbers for that year (using Equation 3.1). Indicator 2 is formed by calculating the geometric mean (Equation 3.2). No variable is considered here as being more important than another and hence equal weights are applied to the variables. Both composite indicators show that Bulgaria's education system has involved many more human resources over the 12 year period. The input into the system shows a 25 per cent increase by 1972 over the corresponding input from the Bulgarian society in 1960. Over the 1965-72 period, this increase has been at a fairly constant rate. Prior to that, the rate was much more rapid due particularly to the expansion of the third-level. The arithmetic and the geometric means in this example are very similar in value because there are neither extreme values in nor major fluctuations across the set of component values.

Index numbers of this straightforward type can also be used for limited norm-referenced comparisons. For example, on the same indicator as used for Bulgaria, the corresponding index numbers[1] for

A. Basic values

Year	1st-level enrolment rates		2nd-level enrolment rates		3rd-level students per 10,000 population	Teachers per 10,000 population
	male	female	male	female		
1960	94	92	73	45	78	97
1965	97	97	73	45	122	116
1970	98	96	83	59	118	118
1972	98	98	85	63	134	120

B. Index numbers

Year	1st-level enrolment rates		2nd-level enrolment rates		3rd-level students per 10,000 population	Teachers per 10,000 population	Indicator	
	male	female	male	female			1	2
1960	100	100	100	100	100	100	100	100
1965	103	105	100	100	156	120	114	113
1970	104	104	114	131	151	122	121	120
1972	104	107	116	140	172	124	127	125

Source: Calculated from Unesco *Statistical Yearbook* 1974

Table 3.9: *Basic values and index numbers for indicator measuring human resource involvement in the provision of education – Bulgaria 1960 to 1972*

France are:

for 1960 100
1965 110
1970 138
1972 135.

The same value of 100 for both countries in 1960 in no way implies the two countries were at equal levels on the indicator in that year. Instead the values establish a corresponding base point so that the amount of change can be ascertained. The 1972 value of 135 for France shows a 35 per cent increase in the indicator value had occurred compared to the 27 per cent increase for Bulgaria.

A comparison between the two education systems in this way is perfectly proper. Indeed it has the additional advantage of lending a second basis for interpreting values for a particular system. Very often, however, a comparison involving the introduction of a second education system is not considered when index numbers are used.

Differential weights could have been applied to the component variables of the above index if an implicit assumption has been made

---

1. The method of combination is the arithmetic mean.

that the value or importance of each variable differed. Different levels of the system, for example, have different priorities attached to them in political policy statements or in development plans. Where such differences are identifiable in this way, different weights should be used in forming composite indicators.

An education cost index

One application of the concept of using differential weightings to form a self-referenced indicator from index numbers can be demonstrated by calculating an education cost index. This index measures, over a certain time period, change in the price of offering education in a single country. It does not measure the total cost of education because this is subject to variation through both increased enrolments and increased costs. The education cost index measures how much is spent relative to a base year in providing the education system, regardless of changes in enrolments. In such a situation, it is not appropriate to examine total expenditure across all levels both because of the very different costs of providing each level and because of the differences in numbers enrolled.

An example demonstrates the formation and implication of an education cost index. Table 3.10 records the per pupil costs and the enrolments in the 1965-72 period for Kenya. Merely ratioing each cost figure to the 1965 base date and averaging (or ratioing total expenditures at each level across years) does not reflect the very large differences between the numbers of students receiving education at each level. In Kenya for 1965, the third-level per student cost is 78 times that for the first-level, but the first-level enrolment is 325 times that for third-level.

Year	Per pupil costs (shillings)			Cost index	Product index
	1st-level	2nd-level	3rd-level		
1965	158	1610	12392	100	100
1970	144	992	6100	77	196
1972	241	1024	8959	114	235

Year	Enrolments		
	1st-level	2nd-level	3rd-level
1965	1010889	54069	3109
1970	1427589	135740	9415
1972	1675919	172602	10468

Source: Calculated from Unesco *Statistical Yearbook* 1974

Table 3.10: *Education cost and product indicators for Kenya – 1965 to 1972*

Two ways to calculate the education cost index to accommodate the enrolment differences involve using the Laspeyres formula or the Paasche formula. The details of these will be discussed shortly. At this stage it is preferable to show the usefulness of the former in defining the cost index. The calculated index values are also reported in Table 3.10. They show that unit costs of Kenyan education over all pupils being educated at any of the levels fell substantially over the 1965-70 period. They then subsequently rose over the 1970-72 period to such an extent that by 1972 relative unit costs were overall above those prevailing in 1965. This increase is mainly attributable to the increase in the first-level per pupil costs. Hence, to educate the same number of students in 1972 as had been educated at each level in 1965 would cost the Kenyan authorities 14 per cent more, while the 1970 figures indicate that costs would be 23 per cent less. An education cost index such as is recommended here thus allows planners to evaluate the impact of their plans for greater efficiency in the delivery of education by controlling the natural and overlapping differences and increases of the actual total enrolments and the actual total expenditures which confuse any direct conclusion about efficiency and relative costs. Certain changes in unit costs at each level can be seen in the data reported in Table 3.10 and used in computing the cost index. The advantage of the indicator is that it allows the importance of these changes in unit costs to be assessed with respect to the number of pupils being educated at each level in a given year.

The Laspeyres formula for the education cost index for levels 1 to 3 can be expressed as follows:

$$CI = \frac{\sum_{i=1}^{3} E_{ib}.C_{it}}{\sum_{i=1}^{3} E_{ib}.C_{ib}} \qquad \textit{(Equation 3.12)}$$

where CI = education cost index in year t

E_{ib} = enrolment at level i in the base year b

$.C_{it}$ = per student cost for level i in year t.

When per pupil costs do not change from those operating in the base year or they change at the different levels by amounts proportional to the base year enrolments at each level, the index number which is calculated is 100. Interpretation is, as has been shown, identical to that for normal index numbers. The use of the weighting factor of enrolments in the base year (E_{ib}) is sometimes criticized. It does introduce the weakness that when there is a fall in demand, the Laspeyres formula will overstate price increases. Such a situation does not happen often in education systems and so the Laspeyres formula is adopted here.

The Paasche formula attempts to overcome this problem by using as weights, the enrolments in the most recent year. That is the E_{ib} relate to the last year in the time series being analysed and not to the first year. This modification does not really solve the problem, however, because when there is a fall in demand, the Paasche formula underestimates the increase. Additionally, up-to-date information is often difficult to obtain for some education system variables to allow the calculation of a complete Paasche series for each of a number of indicators. It also comments in a backward-looking rather than a forward-looking way which is conceptually different from the view preferred by most planners. When the situation of falling demand is very marked, a chained Laspeyres formula can be used. This involves recalculating the weights by averaging enrolments over consecutive years at each step of the calculation. The computing formula[1] for this alternative is reported in Fowler (1970, 17).

An education product index

The Laspeyres formula also allows another potentially useful indicator to be calculated to comment on education system development. This index, which is by nature analogous to the education cost index just described, is the education product index. It seeks to determine how much more it costs, in terms of prices established in a base year, to operate an education system, given the change in the numbers enrolled at each level. Either the Laspeyres or the Paasche formula can again be used with the modification that enrolments not costs are now allowed to change. Costs in the base year or the final year are now used as the weights.[2] Table 3.10 shows that by 1970, a 96 per cent rise in the educational budget for Kenya would be expected if 1965 unit costs had been maintained. By 1972, a 135 per cent increase over the 1965 expenditure would be expected. These values, compared to the rises which actually did occur (and which are subject to inflationary pressures) reflect the greater efficiency with which Kenyan education was offered in 1970 and 1972.

The education product index provides values which should form the basis for projections by mathematical models. The index values estimate likely total costs at current prices if present methods of providing

---

[1] Further detailed discussion of the Laspeyres and the Paasche formulae and of their derivative formulae can be obtained in Crowe (1965) and Parsons (1974).

[2] The Laspeyres formula for the education product index (using the same symbols as for Equation 3.12) is:

$$PI = \frac{\sum_{i=1}^{3} E_{it}.C_{ib}}{\sum_{i=1}^{3} E_{ib}.C_{ib}}$$

education are continued. The index could therefore be used as a dependent variable in a projection equation with time serving as the independent variable (if no more suitable alternative could be nominated). Possible future values for the product index could then be estimated. These estimates would be independent of errors involved in projecting enrolments. Comparison of values projected for the product index with those projected by enrolment projections supplemented by cost estimates might provide a more realistic estimate than has been provided by most mathematical models to date.

Both the cost index and the product index offer planners informative but very different comment on education system development. The former is designed to answer the question: 'Given changes in the unit costs of the system, how much more (or less) would it cost to educate the same number of students who were enrolled in a nominated base year?' The latter index is designed to answer the question: 'If unit costs remain as they were in a given year, how much would the education budget be expected to change because of the increase in student numbers?' The Laspeyres formula perhaps offers a slightly better comment on the changes than does the Paasche formula. Either however can be used effectively to compute index numbers from which to answer these most important planning questions.

Further applications of the Laspeyres and Paasche formulae

Although both the Laspeyres and the Paasche formulae have been traditionally applied in the context of prices and quantities, this is not the only use to which they should be put. A number of other situations are equally appropriate in which both quantities and values change. One of these contexts is the degree of satisfaction a society has with the education system. This is classifiable as an output indicator. To demonstrate the way in which index numbers can be calculated here, data from the Gallup polls of public attitudes towards US education are used (see Gallup, 1970-76). Over the seven year period covered by the second to the eighth polls, Americans in the Gallup samples consistently identified a core of seven major problems confronting the public schools. These issues are listed in Table 3.11. Other problems identified in some years but not in others are excluded from this list. Also reported in Table 3.11 is the percentage of respondents in each year who cite each of the seven issues as being problems. All of these seven variables can be combined into an index to measure the degree to which the American community considers these major problems have been alleviated over the 1970-76 period. Merely summing the percentages is confounded by the possibility of multiple responses. It also equates the rise in the percentage for a minor problem with the corresponding fall in the percentage of respondents citing a major problem. Index numbers calculated by using the Laspeyres formula are reported in Table 3.11

with 1970 used as the base year.[1] The greater the decrease in the percentage for the problem ranked in 1970 as being more important, the lower the value of the index. Not weighting the problems by the 1970 ranks equates the relative importance of the problems which is not reasonable in view of the differences in the percentages.

Problem	Percentage of respondents citing problem each year						
	1970	1971	1972	1973	1974	1975	1976
Lack of discipline	18	14	23	22	23	23	22
Integration	17	21	18	18	16	15	15
Proper finances	17	23	19	16	13	14	14
Hiring 'good' teachers	12	11	14	13	11	11	11
Drugs	11	12	4	10	13	9	11
Poor curriculum	6	3	5	7	3	5	14
Lack parent interest	3	4	6	4	2	2	5
Index value	100	105	110	108	101	99	103

Source: Gallup (1970-76)

Table 3.11: *Indicator measuring change in perception of central educational problems in the USA – 1970 to 1976*

The values for the 'perception of central education problems' indicator show that no diminution of the importance of the problems is evident. Slight fluctuations occur in some years but overall the satisfaction value is about the same in 1976 as it was in 1970. To be able to determine this by examining just one indicator value provides a ready summary of the relatively diverse details which are combined to form the composite indicator.

The Laspeyres formula in particular should contribute quite significantly to the development of self-referenced education system indicators. It has many advantages which planners should exploit wherever possible. One of these is that it condenses an amount of information spanning many years and reports this as a series of easily comprehendable index numbers. These numbers are familiar to almost everyone and consequently acceptance of such an indicator is facilitated. The direct incorporation of an appropriate and educationally defensible weighting component further enhances its use for measuring and monitoring change within a system over a given time period. The three examples of using the Laspeyres formula presented here demonstrate part of its versatility. Planners and researchers are encouraged to extend this range of applications.

---

[1] The 'percentage respondents' is the quantity being indexed while the 1970 ranks (from 8 = most important) provide the weights. The index thus regards discipline, integration and financing as being more important problems than lack of parental interest or school board policies.

Indicators calculated from promotion and repetition rates

Although indicators measuring absolute quantities should generally not be calculated, there are some indicators which do proffer insight into the functioning of an education system when examining it in a self-referenced way. These indicators are therefore exceptions to the general rule established in Section 1.3. A number of such indicators providing useful comment on a system's operation can be calculated from the matrix of transition ratios. This matrix summarizes the transition rates between and within grades – that is the promotion and repetition rates. Three indicators are particularly useful. Two of these are process indicators and one is an output indicator. The process indicators[1] are:

- ☐ the expected number of years of education to be received by an entrant
- ☐ the expected number of grades in a level to be completed by an entrant.

One major concept underlying both of these indicators can be expressed as follows. For those presently enrolled in grade 1 of a given level, the education system delivers one year of education. Only those who drop out or who are promoted to the next grade, however, can be classified as really having completed that grade. Thus if there are no repeaters, the average education delivered to each student is one year. If there are repeaters (ie students intending to re-do that year of education) the average education received by each student who was originally in that grade must be greater than one year. The expected number of years of education to be provided in any one grade can thus be expressed as follows:

$$\text{Years of education in grade i} = \frac{1}{1 - \text{repetition rate}} = \frac{1}{1 - r_{ii}} \qquad \textit{(Equation 3.13)}$$

As all entrants do not reach all grades, the value for each grade as calculated by Equation 3.13 must be weighted by the proportion of entrants who reach each grade. This value is determined by calculating the number of students who leave each grade and who continue their studies. This quantity is called a continuation coefficient and it can be expressed as:

$$\text{continuation coefficient} = \frac{\text{proportion of students promoted from grade i to grade j}}{\text{proportion of students not repeating grade i}} = \frac{p_{ij}}{1 - r_{ii}} \qquad \textit{(Equation 3.14)}$$

---

[1] The indicators can be re-defined to measure respectively the years of education expected to be received or the grades of education expected to be completed by those students presently enrolled in any grade.

The expected number of years of grade 2 education received by those students who enter grade 1 can now be calculated. It is:

$$\begin{array}{lclcl} \text{expected number of years} & & \text{expected number} & & \text{proportion of} \\ \text{of grade 2 education} & = & \text{of years of grade 2} & \times & \text{entrants} \\ \text{received by grade 1} & & \text{education} & & \text{reaching} \\ \text{entrants} & & & & \text{grade 2} \\ & = & \dfrac{1}{1 - r_{22}} & \times & \dfrac{p_{12}}{1 - r_{11}} \end{array}$$

Generalizing this concept to all grades in a level and summing the terms for each grade to find the total expected years results in the following formula:

$$\begin{array}{l} \text{expected number of years} \\ \text{of education to be} \\ \text{received by an entrant} \end{array} = \frac{1}{(1 - r_{11})} + \frac{p_{12}}{(1-r_{11})(1-r_{22})} + \frac{p_{12}\,p_{23}}{(1-r_{11})(1-r_{22})(1-r_{33})} + \ldots$$

(Equation 3.15)

The formula continues to develop as in the model evolving above. For six grades, the last term to be summed would be:

$$\frac{p_{12}\,p_{23}\,p_{34}\,p_{45}\,p_{56}}{(1-r_{11})\,(1-r_{22})\,(1-r_{33})\,(1-r_{44})\,(1-r_{55})\,(1-r_{66})}$$

Substituting into Equation 3.15 can be time-consuming. When moderately sophisticated computing facilities are available, it is more efficient to use elementary matrix algebra. This can simplify the calculations considerably. (Most programmable desk calculators can be used to do such a task.) The years of education are calculated[1] by:

$$YE = (I - T)^{-1}$$

(Equation 3.16)

where T = transition matrix containing promotion and repetition rates for each grade in a level.

The number in row i and column j of the resulting YE matrix is the number of years of education delivered in grade j to those students presently enrolled in grade i. The value for the indicator being defined here, therefore, is the sum of the elements in the first row. It is the sum of the years of education to be received in each grade by students now in grade 1.

An example demonstrates the way in which this indicator can be estimated and used. Table 3.12 reports the promotion and repetition rates for 1969-70 and 1973-74 for Saudi Arabia for males enrolled at the first-level. Each of the columns in this table forms a transition matrix (T). This matrix can be manipulated as described by Equation 3.16 or the individual values can be substituted into Equation 3.15.

---

[1] Full details of the calculations and the mathematics involved for the general case can be found in Thonstad (1969) and Johnstone and Tobing (1976).

Transition	1969-70	1973-74
Repeat grade 1	0.307	0.207
Grade 1 to grade 2	0.627	0.732
Repeat grade 2	0.223	0.156
Grade 2 to grade 3	0.737	0.831
Repeat grade 3	0.227	0.167
Grade 3 to grade 4	0.749	0.826
Repeat grade 4	0.325	0.269
Grade 4 to grade 5	0.630	0.692
Repeat grade 5	0.264	0.177
Grade 5 to grade 6	0.683	0.796
Repeat grade 6	0.253	0.144
Graduate from grade 6	0.700	0.800

Source: Saudi Arabian ministry of education *Annual Report*

Table 3.12: *Promotion and repetition rates for first-level education – males, Saudi Arabia, 1969-70 and 1973-74*

The results of transforming both transition matrices are reported in Table 3.13. To males who entered first grade in 1969, the number of years of education, on the average, that can be expected to be delivered are calculated as:

- ☐ in first grade, 1.443 years ($= \frac{1}{1 - 0.307}$)
- ☐ in second grade, 1.164 years ($\frac{1}{1 - 0.307} \times \frac{0.627}{1 - 0.223}$)

For males who were enrolled in the third grade in 1973, the corresponding values for the expected number of years of education to be received are:

- ☐ of third grade education, 1.201 years ($= \frac{1}{1 - 0.167}$)
- ☐ of fourth grade education, 1.357 years ($= \frac{1}{1 - 0.167} \times \frac{0.826}{1 - 0.269}$)

The values in the top line of each of the matrices in Table 3.13 correspond to the terms defined in Equation 3.15. Summing these values therefore provides the estimate that an average of 6.94 years of first-level education must be provided for males who entered the first grade of the Saudi Arabian education system in 1969. This value compares to the average of 6.68 years of first-level education which must be provided to males who entered the Saudi system in 1973. A decrease of about one quarter of a year thus appears to have been

a) 1969-70		**Years of education delivered in grade**						
		1	**2**	**3**	**4**	**5**	**6**	**Total**
Grade in which student begins	1	1.443*	1.164*	1.104	1.225	1.049	0.959	6.944*
	2	0.000	1.287	1.220	1.354	1.159	1.060	6.080
	3	0.000	0.000	1.294	1.436	1.229	1.124	5.083
	4	0.000	0.000	0.000	1.482	1.268	1.160	3.910
	5	0.000	0.000	0.000	0.000	1.359	1.242	2.601
	6	0.000	0.000	0.000	0.000	0.000	1.339	1.339

b) 1973-74		**Years of education delivered in grade**						
		1	**2**	**3**	**4**	**5**	**6**	**Total**
Grade in which student begins	1	1.261	1.094	1.091	1.233	1.037	0.964	6.680*
	2	0.000	1.185	1.182	1.336	1.123	1.044	5.870
	3	0.000	0.000	1.201*	1.357*	1.141	1.061	4.760
	4	0.000	0.000	0.000	1.368	1.150	1.070	3.588
	5	0.000	0.000	0.000	0.000	1.215	1.130	2.345
	6	0.000	0.000	0.000	0.000	0.000	1.168	1.168

* These values are explained in the text.

Table 3.13: *Estimated years of education to be delivered at the first-level – Saudi Arabia, 1969-70 and 1973-74*

brought about over the period being considered in the length of time a Saudi entrant can expect to spend studying at the first-level.

In the context described, the calculated indicator values can be directly compared either with each other or with the formal length of the level of education being analysed. Where different numbers of years of education are formally provided, dividing the calculated value by the formal number of years offered forms an appropriate basis for either self-referenced or norm-referenced comparisons.

The second process indicator noted above as being derived from a transition matrix is the number of grades expected to be completed by an entrant. This is conceptually different from the indicator just discussed although, in a highly developed education system, the values estimated for the two indicators should be about the same. While the previous indicator measures the length of time a student can be expected to spend in a level, this second indicator measures how many grades that student can be expected to complete in that period. Hence the first indicator measures 'how long' while this indicator measures 'how far'.

The formula for the number of grades expected to be completed by a grade 1 entrant is:

$$\text{number of grades expected to be completed by an entrant} = \frac{p_{12}}{(1-r_{11})} + \frac{p_{12}\,p_{23}}{(1-r_{11})(1-r_{22})} + \frac{p_{12}\,p_{23}\,p_{34}}{(1-r_{11})(1-r_{22})(1-r_{23})} + \dots$$

(Equation 3.17)

where the symbols are as in the previous example.

This formula extends the previous one summarized by Equation 3.15 by modifying each term by the corresponding promotion rate. This is done because those who drop out, complete their education at that grade level. Only those who continue in the system (ie who are promoted) increase the expected number of grades to be completed. In calculating values for this indicator by using Equation 3.17, values estimated for the continuation ratios can be used from the previous indicator calculations or each of the terms in Equation 3.15 modified simply by multiplying them by the appropriate promotion rates.

To estimate values for this indicator using matrix algebra, a new matrix must be formed. It consists of only the promotion (or graduation) rates from each grade. It is therefore of size (g x 1) where g is the number of grades in that level. This matrix – call it P – is then pre-multiplied by the previously formed matrix. Hence:

$$GE = (1 - T)^{-1} . P \qquad \textit{(Equation 3.18)}$$

The elements of this GE matrix give the number of grades expected to be completed by students starting from each grade in a level.

The calculation of the expected number of grades to be completed can be demonstrated by extending the example for Saudi Arabia introduced for the previous indicator. The value for the expected number of years an entrant spends in each grade is given by each value in the top line of the matrices in Table 3.13. These values, which were the components of Equation 3.15, must each be multiplied by the appropriate promotion rate, as is described by Equation 3.17, and then summed. Hence:

$$\begin{array}{ll} \text{expected number of grades (1969 entrants)} & = 1.443 \times 0.627 + 1.164 \times 0.737 + \ldots \\ & \quad \ldots + 0.959 \times 0.700 \\ & = 4.75 \end{array}$$

For 1973 entrants, the calculated value is 5.18 grades. Hence, the average level of attainment of Saudi first-level males appears to have risen by about one third of a grade over the 1969 to 1973 period.

A third indicator noted above as being derived from a transition matrix is classifiable as an output indicator. It is the proportion of entrants who finally graduate. The formula for the calculation of values is:

$$\text{proportion of entrants who graduate} = \frac{p_{12}}{1-r_{11}} \cdot \frac{p_{23}}{1-r_{22}} \cdot \frac{p_{34}}{1-r_{33}} \cdot \ldots \cdot \frac{p_g}{1-r_{gg}} \qquad \textit{(Equation 3.19)}$$

where the symbols are as before and 'g' is the last grade in the level.

The conceptual basis for this formula is again the continuation ratio defined above – ie the proportion of students who leave each grade and continue their studies. Multiplying the continuation ratios together for each grade hence produces the proportion of entrants who finally graduate.

For the data in Table 3.12, the following proportions can be calculated:

$$\text{for 1969-70 } I = \frac{0.627}{1-0.307} \cdot \frac{0.737}{1-0.223} \cdot \frac{0.749}{1-0.227} \cdots \frac{0.70}{1-0.253}$$

$$= 0.675$$

$$\text{for 1973-74 } I = \frac{0.732}{1-0.207} \cdot \frac{0.831}{1-0.156} \cdot \frac{0.826}{1-0.167} \cdots \frac{0.800}{1-0.144}$$

$$= 0.771$$

Hence, approximately 67 per cent of males who entered Saudi elementary schools in 1969 could have been expected to graduate from that level. For 1973 entrants, the corresponding value is 77 per cent. The holding power of Saudi elementary schools has thus increased dramatically over a very short period.

Using the matrix algebra, the calculation is less elegant and more involved. The transition matrix (T) really represents what the distribution of enrolments is like in a given level of an education system after one year. Multiplying this matrix by itself to form T^2 thus shows the distribution of students after two years. In general, T^n represents the distribution of students after n years.

The value in row 1 and the last column in each of the sequence of T^n matrices formed by multiplication represents the proportion of grade 1 entrants who reach the last grade of that level in the specified time period. Adding these values from consecutive calculations of T^n forms the total proportion of entrants who ever reach the last grade. For a 3 grade level, powers to 5 are generally sufficient while for a 6 grade level, powers to 8 are adequate. Multiplying the total sum by the graduation rate gives the graduates as a fraction of the entrants.

Distributive self-referenced indicators

To measure the distribution of a quantity throughout a single system to facilitate self-referenced comparison, both the Gini coefficient and the C measure of concentration can be used. Continuing the example for the Gini coefficient begun in Table 3.7 allows the progress in Nicaragua to be traced from 1965 to 1972. Diagram 3.3 depicts the Lorenz curves for the distribution of enrolments at the first-level across each grade. The Gini coefficients calculated for these curves are:

for 1965 $G = 0.41$
1970 $G = 0.34$
1972 $G = 0.31$

These values show there has been a definite and gradual increase in the equality of the distribution of enrolments across grades. This is also shown in the diagram. Perhaps this increase is due to an increase in the quality of the first-level provision leading to a corresponding decrease in drop-out and repetition rates. If this is so, planners, administrators and teachers in Nicaragua are to be commended. Identification of the precise reason for the change in the Gini coefficient would be a useful exercise for planners in that country.

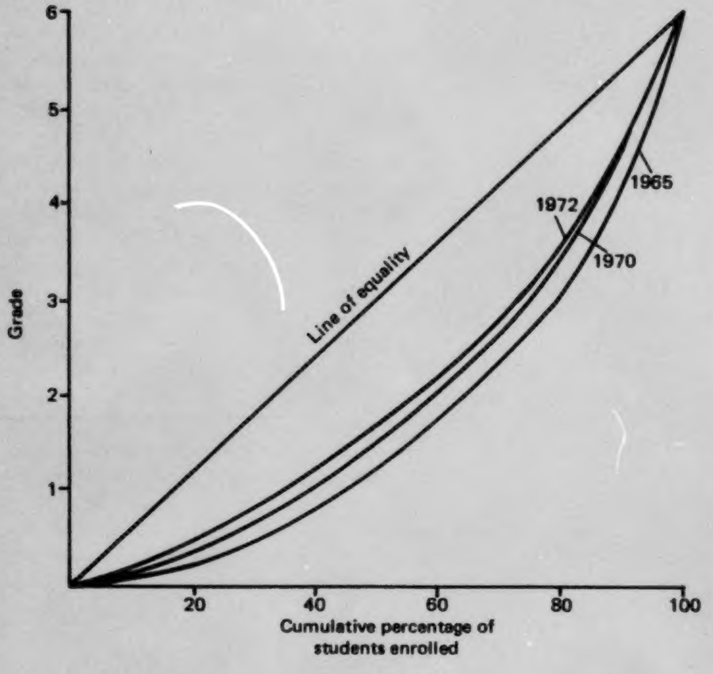


Diagram 3.3: *Lorenz curves measuring inequality of distribution of enrolments across grades at the first-level — Nicaragua 1965-72*

The application of the C index as a self-referenced indicator parallels this extension of the Gini coefficient. Provided the cautions discussed previously are observed, no problems should be encountered by applying the formula for the C index to values spanning a number of years for a single education system.

3.4 Criterion-referenced indicator formation

Unlike the situation described for norm- and self-referenced indicators, special indicators allowing criterion referenced interpretation are not normally developed. This is not to deny that criterion-referencing is often an integral part of planning – especially as part of the monitoring process. Instead it admits that indicators developed in any of the ways already described can be used to determine whether an education system is operating at a level deemed to be satisfactory. When a concurrent evaluation of a number of systems is being undertaken in this mode of comparison, a critical value to indicate an acceptable level of development can easily be imposed on the distribution of the values for a particular indicator or set of indicators. When a single system is being examined over a period, index numbers or some other appropriate format can be mapped to determine either that progress is being made towards a goal or that an acceptable level of system functioning is being maintained.

There are two indicators, however, which do represent the main ways in which criterion-referenced indicators have been developed. The first of these is that proposed by Harbison *et al* (1970). They propose the concept of an 'ideal country'. This involves defining the absolute level desired for an education system on each of a number of variables. Measures of the 'distance' any particular system is away from each of these ideal values are then calculated and summed in a similar way to that described in Section 10.2. If a system has a large indicator value for its distance away from the ideal, then it is substantially different from the criterion; if it has a small value for the indicator, it is very close to the ideal. Such an indicator does not however make a substantial comment on the development of a system except in very special circumstances. Further discussion on this point is therefore deferred to Section 10.2.

The second type of criterion-referenced indicator is proposed by Correa (1975). It measures the degree to which plans have been successfully implemented for a particular education system. Correa defines the following index:

$$I = \frac{1}{D} \sum_{j=1}^{2} \sum_{i=1}^{N} \frac{(A_{ij} - P_{ij})^2 \times D_{ij}}{P_{ij}} \qquad \textit{(Equation 3.20)}$$

where P_{ij} = target rates of growth across the ith planning period for educational level j

A_{ij} = achieved rates for growth across the ith planning period for educational level j

D_{ij} = duration of the ith plan for the jth educational level

$D = \sum_i \sum_j D_{ij}$.

The higher the value of I, the less successful has been the implementation of plans in that system. So, for example, Correa calculates a value for Argentina of 0.67. This shows that Argentina has been fairly successful in achieving its planned targets. In contrast, Bolivia, with a value of 4.57, and Uruguay, with a value of 12.63, have been less successful. This implementation index thus provides a good basis for planners to assess the extent of the effect their plans have on an education system.

In non-educational settings, indicators are often constructed to relate the level of something being measured to a criterion. One field in which considerable indicator developmental work has taken place is the measurement of environmental quality. Here indicators have been proposed to measure water quality, air quality, urban environment quality and pollutant burdens in humans. Of these, air quality indices in particular have been extensively refined. In many areas, especially in the US, the daily values of an air quality index (AQI) are reported, together with temperature and humidity, as part of the usual weather report. The values on the AQI are related to a 'healthiness scale' to enhance their interpretability. Criterion values must therefore be adopted to form a basis for the degree of purity (or impurity) of the amount of pollutants in the air.

A typical criterion-referenced air quality index is that proposed by Babcock and Nagda (1972). They utilize five of the six air pollutants for which standards have been laid down by the US Environmental Protection Agency. These pollutants are: carbon monoxide, sulphur oxides, nitrogen dioxide, oxidants and particulate matter. Their AQI is then defined as:

$$I = \left[5.7 \sum_{i=1}^{5} \frac{C_i}{S_i} \right]^{1.37} \qquad \textit{(Equation 3.21)}$$

where C_i = measured concentration of pollutant i
S_i = EPA standard for pollutant i.

The coefficient of 5.7 and the exponent of 1.37 are scaling constants which ensure values calculated for the indicator lie between 10 and 100 when pollutant levels are less than their limits. The value for the exponent is also claimed to have some empirical support. The form of the index is based on 'an intuitive feeling that an increment of pollutant added to clean air is less harmful than the same increment added to already-polluted air' (1972, 190).

The difficulty in applying such criterion-referenced indicators as this to education is that few agreed to standards of performance exist. Pollutants etc are comparatively easy to measure and their effects can be determined under proper, scientifically controlled conditions. For education and the provision of systems for its delivery, it is not so easy to establish critical standards. Achievement levels, a certain minimum number of grades of schooling before achieving functional literacy, classroom space and class sizes are all variables for which minimum standards have been considered to exist. No sound empirical basis for the values proposed has however been consistently established. For cross-national comparisons of education systems, use of these assumed standards would be invalid. For use with systems aggregated to a regional level only, indicators using a formula comparable to that described by Equation 3.21 might be able to be adopted. In such a setting, conditions are more homogeneous. They are also subject to similar or the same policies. Assumptions would, of course, have to be made about the scaling constants. Although standards are of concern to all educational planners, few opportunities presently exist for the measurement of education systems related to any criteria or standards of provision and performance.

3.5 The problem of interchangeability

When indicators are formed by either theoretical or empirical definition or both, two or more indicators can be sometimes considered to measure basically the same education system characteristic. When this happens and one indicator only must be chosen, the problem arises as to which indicator is most appropriate. This problem can be referred to as the problem of the 'interchangeability of indicators'. No single solution to the problem exists, although three alternatives to selecting a single indicator may be followed.

One method of ascertaining which indicator might be best involves examining all similar indicators to determine if they are equal in sophistication as well as being mathematically similar in their formation. Do they, for example, combine variables measured on scales of similar measurement sophistication? Interval and ratio scales, when appropriate to use, allow much more critical comment and measurement of education system characteristics than do nominal or ordinal scales. Data availability plus data alleged or known to contain at most only small errors of measurement are also to be preferred in any indicator finally chosen. The suitability of adding or multiplying together component variables to form an indicator also has to be decided from a conceptual point of view. Through comparing and contrasting parallel indicators in this way, some differentiation between indicators and a ranking of their order of suitability for a particular purpose can sometimes be established.

A second way to examine similar indicators is to plot values for each indicator against one another across the entities which are to be measured – whether they be nations, provinces or individual schools. Any differences in classification can then be easily identified and a decision made as to which indicator more properly classifies or orders the entities being examined.

Such an exercise is demonstrated in Diagram 3.4. There the two indicators being compared are the original Harbison and Myers composite index (plotted along the horizontal axis) and its alternative form calculated so that the third-level enrolment ratio is five times more important than the second-level enrolment ratio in forming the composite value (plotted along the vertical axis). This alternative index is the one Harbison and Myers originally argue for and is the one referred to earlier in Section 3.2. Both indicators therefore measure the extent of involvement of a community in second- and third-level education.

Along the horizontal axis of Diagram 3.4 are marked the original divisions of countries as determined by Harbison and Myers while along the vertical axis divisions forming country groups of the same size are marked. Countries falling outside similar classifications can now be easily identified. For example, Yugoslavia and Czechoslovakia are classified by Harbison and Myers as being in level III while the alternative index puts them into level IV. Denmark and West Germany are displaced from level IV to level III. Similarly Lebanon and Jamaica become level III countries under the alternative index at the expense of Thailand and India which move down to become level II countries.

Countries on the borderline of a level can also be re-examined. Argentina, for example, is only marginally in level IV when the old index is used. It is however clearly a level IV country under the alternative index. Mexico is confirmed as being in level III. Iraq and Peru pose an interesting situation because of their low level enrolment ratios in 1960. The original Harbison and Myers index puts them right at the border between level II and level III because of their relatively higher second-level enrolment ratios counteracting the low third-level ratios. The alternative index however places both countries at the border between level II and level I – a considerable change of status.

On the basis of all of these modifications and qualifications to the component countries of each level as formed by each indicator, guidance is given for a decision to adopt one rather than another indicator.[1] This guidance in no way ensures that a valid or 'the correct' indicator will be selected. The approach does however demonstrate what the implications are for selecting one rather than another indicator.

An alternative to this approach involving the plotting of values can also be used. It involves calculating regression lines between the various alternative indicators. Distances from the several lines can then be estimated and consideration focuses on those entities which are furthest away from particular regression lines (ie which have the largest residual

---

[1] Some of the possible modifications noted above are argued *a posteriori* by Harbison and Myers as being possible weaknesses of their index (see, for example, pp 74 and 145). In this situation no recommendation for one index rather than another can be made because no definite use for either indicator is envisaged.

134

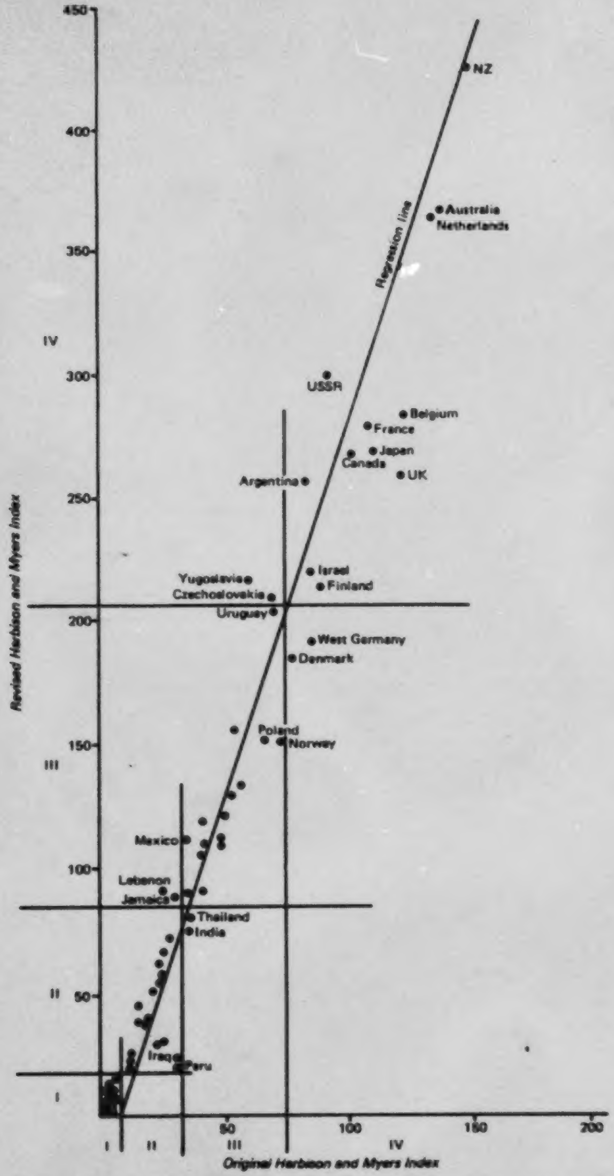


Diagram 3.4: *Two forms of an indicator measuring involvement in second- and third-level education*

between their actual and predicted values). In Diagram 3.4, for example, the line of regression of the old index on to the alternative index is drawn. Its equation is:

old index = 5.98 + 0.33 x new index

The regression line of the new index on to the old could also have been calculated and plotted. Countries such as are noted above are immediately identified by their residuals as those which must be given close consideration before one indicator in particular is chosen to represent the concept being measured.

A third way to determine the extent to which indicators are interchangeable is to assess their inter-correlations. This is mathematically allied to the regression approach just outlined. Here, however, the extent of the inter-relationship is assessed while the regression approach aims to examine the magnitude of the extreme residuals. When two indicators inter-correlate very highly, some degree of interchangeability is indicated. A high value thus identifies a situation in which one indicator rather than another should be used. When low inter-correlations exist, a degree of independence – at least in the statistical sense – is revealed between the indicators. In this latter situation, two different dimensions might really be present so that both indicators might need to be reported.

In the Harbison and Myers example discussed above, the inter-correlation between the indicators is very high – about 0.98. Such a value indicates that for the majority of countries, either index could be adopted.

Generally it is preferable to use all three methods to assess the interchangeability of members within a set of indicators identified as measuring education system characteristics in the same general conceptual area. A set defined in the input-resource category or the output-satisfaction category, for example, might be analysed to identify those indicators reducing the parsimony of that set. In other situations, the task of collecting sufficient data to allow the statistical procedures to be performed is too time consuming.

Conceptual and mathematical analysis are the only methods by which the degree of interchangeability can be established.

3.6 Résumé

Planners and researchers must decide on every aspect of the process in their development of theoretically defined indicators. They must carefully select the variables to be combined, the methods for the combination and the weights to be used. As the discussion above reveals, there are very many different ways to facilitate the construction of theoretically defined indicators. The main methods discussed for norm-referenced and for self-referenced indicators are summarized in

Table 3.14. Distinctions must be carefully drawn between whether distributive or non-distributive indicators are to be formed as well as among the kinds of scales on which variables are to be measured. As Table 3.14 shows, some indicators can be used in more than one situation. Generally, however, different methods need to be employed for different purposes. When forming a basic set of indicators, it is not at all necessary to use indicators of just one type. A variety of education system characteristics must be measured in any situation and hence a variety of indicator types should be used as it is appropriate.

Comparison intended	Measurement intended	Type of scale	
		Nominal Ordinal	Interval Ratio
Norm-referenced	Non-distributive	Restricted — summation and/or Likert scale analysis	Standard scores
	Distributive	C measure of concentration	Standard deviation Coefficient of variation Gini coefficient
Self-referenced	Non-distributive	Laspeyres and Paasche formula	Index numbers Special formulae using promotion and repetition rates
	Distributive	C measure of concentration	Gini coefficient

Table 3.14: *Summary of possible methods for developing theoretically defined indicators*

Chapter 4

The Empirical Definition and Development of Indicators

The use of empirical methods to define and develop indicators of education systems contrasts markedly with the theoretical approach discussed in the previous chapter. Here the planner or researcher who has to develop indicators has minimum involvement in determining solutions for the three main problems involved. Certainly some decisions must be made in relation to each problem and these, to some extent at least, do have an effect on the type of indicator produced. However, the most important result — that of determining the weights — is accomplished solely by the technique employed.[1] It is outside of the control of the developer and this aspect in particular differentiates between the empirical and the theoretical definition of indicators.

In this chapter, some of the more important methods of empirical definition are outlined. The methods appropriate for ratio and interval scaled variables are presented first, followed by those for ordinally and nominally scaled variables. Only non-distributive indicator formation is discussed. No appropriate methods for empirically estimating distributive indicators of education systems have yet been developed. Theoretically defined indicators serve this need so well that distributions of characteristics in a population can be very well represented by the variety of formulae available and discussed in the preceding chapter.

Before presenting the empirical methods, three features need to be noted in relation to this general approach to defining indicators. The first is that the characteristics of the indicators depend very much on the entities analysed, the variables included in the analysis and the errors of measurement inherent in the data. Part of this problem has already been discussed in Section 1.6. It is of vital importance that the purpose and the scope of any indicator developed through analysis is identified very precisely before the empirical methods are used. For example, only those entities to which the indicators are to be applied should be included in the analysis. If it is intended to generalize the indicators to entities other than those analysed, then the characteristics of those entities omitted should be very similar if not identical to those analysed. Similarly the set of variables must be selected with care. What sometimes occurs is that indicator developers include all kinds of variables hypothesized to relate to the domain being analysed. It is then hoped that the analytical technique itself will put 'order into the chaos' and 'nice clean indicators' will emerge. This can never happen. All that can be defined is a collection of a sub-set of variables which may or may not even be interpretable. What a developer must do therefore is to identify conceptually similar variables and analyse these separately from other variables which are conceptually dissimilar. An example of this approach is the analysis of a set of input resource variables separately from either a set of input satisfaction variables or from a set of process variables. Interpretation is then facilitated and made very much more meaningful. If it is so desired, the indicator sets can easily be inter-related but this is at a stage beyond that of definition.[1]

---

[1] This statement assumes that the developer has already made particular decisions regarding the analysis and the model on which it will be based. For example, different types of factor analysis will produce different weights. So too will different specifications of variables and their combination in a multiple-regression equation. Once such decisions have been taken however, the technique will determine the relative weights for a given data set.

The error inherent in the measurement of variables must also be examined very carefully. If the error is significant – particularly if it is random error – the estimated relationship between variables will decrease and this in turn biasses the weightings assigned to a particular variable. It might even result in an important variable being deleted from the set used to define a particular indicator or in an unimportant variable being included invalidly. Because of their extreme susceptibility to variation due to any of the three sources noted here, empirically defined indicators must be very cautiously developed and accepted.

A second feature of empirically defining indicators, and one that stems in part from the first, is that the definition should really be demonstrated to be stable both over different samples and over time. If stability cannot be shown over a series of replication analyses, then the results obtained are not really reliable. Limited faith, at best can then be put in the obtained results. This aspect of replication is discussed further in Appendix B. Suffice it to say here, therefore, that more than a single analysis should be undertaken if results are to be accepted as a basis for planning. When data are available for a number of schools or districts and regional education system indicators are being developed, an appropriate but fundamental technique to use involves the division of the total sample into a number of sub-samples.

---

[1] The inter-relating sets of indicators defined for different years and/or for different parts of the education system space can be a most useful exercise in theory development and hypothesis testing. Such an approach is demonstrated in Chapter 9.

When variation occurs in the way particular indicators are defined across these sub-samples, unstable variables should be used very prudently or eliminated from further consideration. Frequently not enough cases are available to allow sub-samples to be analysed. In such circumstances, replication analyses should be performed for at least one other year or data collection period. Comparing results obtained from either of these multiple analysis situations does allow an estimate of the stability of results obtained. If at least moderate stability is established, the defined indicators can then be further developed – if it is appropriate – to form basic education system indicators.

A third feature of empirically defining indicators to measure education system performance is that the techniques are mainly useful for identifying norm-referenced indicators. Empirical methods are largely based on the concept of differentiation among entities. They are therefore really dependent on samples which are at least of moderate size. This restriction can place a very stringent limit on the scope of an analysis aiming to define indicators of education systems empirically. A great deal of developmental work must therefore be transferred to the theoretical definition of indicators especially for self-referenced indicators. Unfortunately when this is done, a number of highly related indicators might be developed. Without empirical evidence, this fact may never be identified and indicators basically measuring the same characteristics might be recommended for inclusion in the same set of core indicators. Empirical methods should thus complement theoretical methods of derivation whenever possible. One kind of approach must not be regarded as being more important than the other. Hence, both should play a part in a field such as education systems where little real development work has yet been undertaken.

4.1 Norm-referenced indicator formation from ratio and interval scales

Factor analysis

By far the most common method of defining indicators from variables is the technique of factor analysis. This is a multivariate technique which aims to reduce a large set of variables down to a much smaller and more manageable set of variable composites. These sets of composites are defined most strongly by a few particular variables from the larger set. For any given composite, the defining variables should be very similar conceptually.

The theory underlying the technique of factor analysis as well as the practical problems involved in its application have been discussed at length in numerous publications. Amongst the most relevant of these to educational planners and researchers are Rummel (1970) and Cattell (1978). Instead of repeating much of their discussion, therefore, an example is now reported to demonstrate how the technique might be used to develop education system indicators.

Allow that a set of 12 variables has been carefully defined to measure specific features of the resources input into an education system. These variables are listed in Table 4.1. Data on each of these variables then had to be collected for every country used in the analysis, there being 66 countries for which this was possible for the year selected, 1960. Hence the factor analysis begins with a complete 66 x 12 matrix of raw data.

	Variable	Loadings	
1.	First-level enrolment ratio – males	0.79	−0.21
2.	First-level enrolment ratio – females	0.81	−0.39
3.	Second-level enrolment ratio – males	0.84	−0.10
4.	Second-level enrolment ratio – females*	0.93	−0.25
5.	Third-level enrolment per 10,000 population*	0.78	−0.22
6.	Teachers per 10,000 population	0.85	−0.24
7.	% GNP spent on education	0.72	0.13
8.	Rate of change in variable 1	−0.42	0.78
9.	Rate of change in variable 2	−0.56	0.68
10.	Rate of change in variable 3	−0.03	0.50
11.	Rate of change in variable 4	0.06	0.45
12.	Rate of change in variable 5	−0.27	0.48
	% of total variance	43.9	17.9
	% of common variance	71.0	29.0

* Variable transformed by a square root transformation before analysis to correct skew in the distribution.

Table 4.1: *Ortbogonally rotated factor matrix of 1960 input variables*

The technique[1] then proceeds to inter-correlate each of the variables across all of the countries. This enables those variables having high inter-relationships between one another to be identified. The factor analytic technique then extracts clusters of highly inter-related variables such that those with the closest relationships are extracted first, then clusters of variables with slightly weaker relationships are extracted, and so on. What results from this process is termed a factor matrix and a typical matrix is reported in Table 4.1 for the 12 defined variables being considered here. Those variables tending to be highly inter-related thus forming a cluster are now identifiable because they are reported as having high loadings on a given pattern. They are therefore referred to as being variables which, taken together, define a particular factor.

An inspection of Table 4.1 demonstrates this situation. If a value of 0.4 is taken (arbitrarily) as being indicative of a variable which defines a factor, then the first factor is defined by variables measuring the input of students into the first-, second- and third-levels and the input of teachers and finance into the system as a whole. The two variables measuring the rate at which the first-level enrolment ratios are changing also have high loadings, although these are preceded by negative signs.

---

[1] There are a number of different kinds of factor analysis. The one being described here is the principal components method.

These nine variables can thus be regarded as defining the first – and most important – way in which the education systems being analysed vary amongst themselves. Hence some systems have high student and teacher input into all levels relative to their populations and they also have high inputs of finance. These systems also have low rates of chahge in the values for the first-level enrolment ratios – presumably because these values are presently close to 100 anyway. Other education systems can be described as being the reverse of this pattern. Taken altogether, these variables seem to be describing education systems in terms of the expansiveness of the involvement in their provision.

Continuing this interpretation, it is possible to interpret the second factor as describing countries which are rapidly increasing the rates at which they enrol students in their education system. This factor thus reflects an emphasis in a system on increasing student involvement in education.

Just as it is possible to report values for each education system for each of the 12 variables being analysed, it is also possible to calculate a composite score for each education system for each factor or pattern identified in the factor analysis. Such a score – termed a factor score – is a type of standardized score as is described in Section 3.2. Hence a system with a high positive value on a given factor is regarded as being high on the composite characteristic or factor which has been identified in the analysis. A system with a high negative score is low on that characteristic. Factor scores around zero indicate countries with about average levels of the characteristic. The two factor scores calculated for each country used in the present analysis are reported in Table 4.2.

Using these scores, it is possible therefore to note that countries such as Belgium, Japan and USA have very high levels of the characteristic termed here 'the expansiveness of the involvement in education systems'. In contrast, countries such as Afghanistan, Ethiopia and Haiti have very low levels. In other words, countries such as Belgium which have high positive factor scores involve very large numbers of people and large amounts of finance in their education systems relative to the totals who could potentially be involved. In contrast, countries with high negative factor scores involve relatively few people and relatively little finance.

For the second factor, countries such as Afghanistan and Ethiopia have high positive values while countries such as Argentina, Czechoslovakia and the Federal Republic of Germany have high negative values. Hence Afghanistan appears to have been emphasizing an increase in the involvement of students in the education system relative to the population of the appropriate age group while Argentina has not.

Country	Factor	number	Country	Factor	number
	1	2		1	2
Afghanistan	−2.11	1.31	Ireland	0.43	−0.49
Albania	−0.50	0.46	Israel	0.98	0.01
Argentina	0.27	−0.86	Italy	0.55	−0.36
Australia	1.00	−0.40	Japan	1.35	−0.02
Austria	1.02	−0.05	Korea, Republic of	−0.22	−0.30
Belgium	1.44	0.02	Lebanon	−0.37	−0.20
Bolivia	−0.73	−0.07	Liberia	−1.72	−0.30
Brazil	−0.26	0.26	Libyan Arab Rep.	−0.82	2.57
Bulgaria	1.33	0.18	Luxembourg	−0.08	−0.56
Burma	−1.26	−0.33	Mexico	−0.77	−0.15
Canada	1.08	−0.06	Nepal	−1.65	1.95
Ceylon	0.20	0.24	Netherlands	1.07	−0.14
Chile	0.16	−0.52	New Zealand	0.89	−0.31
Columbia	−0.72	−0.35	Nicaragua	−1.19	−1.10
Costa Rica	0.17	−0.47	Norway	1.02	−0.21
Cuba	0.29	1.08	Panama	0.25	−0.55
Czechoslovakia	0.36	−0.71	Paraguay	−0.88	−1.90
Denmark	0.59	−0.05	Peru	−0.08	0.32
El Salvador	−0.50	0.53	Philippines	−0.14	−0.89
Ethiopia	−2.22	1.35	Poland	−0.98	−0.01
Finland	0.97	−0.15	Portugal	−0.39	−0.64
France	1.18	−0.24	Rhodesia, Southern	−0.76	0.35
Germany, Fed.Rep.	0.92	−1.07	Romania	0.40	0.67
Greece	0.28	−0.50	Spain	−0.65	−1.00
Guatemala	−1.36	−0.60	Sweden	0.97	−0.12
Haiti	−1.96	−1.05	Thailand	−0.83	−1.01
Honduras	−1.07	−0.18	Turkey	−0.82	0.21
Hungary	0.36	−0.32	United Kingdom	1.24	−0.04
Iceland	0.99	−0.32	Uruguay	0.66	0.05
India	−1.04	0.60	USA	1.50	−0.22
Indonesia	−1.41	−0.84	USSR	1.29	0.14
Iran	−1.01	1.05	Venezuela	0.74	2.24
Iraq	0.38	3.86	Yugoslavia	0.23	0.11

Table 4.2: *Factor scores for each country on the 1960 input dimensions*

These factor scores are therefore exactly like indicator values and they are calculated for each system as follows:

Step 1: estimate standard scores for each variable (using Equation 3.3)

Step 2: multiply the standard scores by the weights calculated by the factor analysis for each variable (see Table 4.1)

Step 3: sum the weighted scores and calculate an average if desired (such as is described by Equation 3.4).

In essence then, the resulting factor score value is just like a value obtained by defining an indicator theoretically except that the weights applied to the component variables are derived empirically.

It is possible to adapt directly the factors as they are empirically defined as indicators in their own right. This is most appropriate when the total set of entities to be dealt with has been analysed. When this is the case, little more needs to be said here. In Chapter 9, an analysis is performed using factor scores because only those education systems originally analysed are to be investigated further there. In some situations, however, indicator values are needed either for entities not included in the original analysis or for the same entities but at a different point in time. In both of these situations, new factor (or indicator) scores must be estimated.

If values comparable to those produced by a factor analysis are required, the value for each variable in the total set must be standardized as described above. This calculation requires mean and standard deviation values to be known for each variable in the original analysis. It also assumes that the 'weighting criterion used is . . . a function of the internal consistency among the component variables' (Blalock, 1960, 389). When other entities are being added to those analysed, this is normally no real problem because their addition, if it is to be valid, should not affect the values for the mean and standard deviation. When values for the same entities are being estimated but for another time period, either constant values for the two descriptive statistics must be assumed (which is to imply that no change has occurred) or new values must be calculated.

It is in calculating new values across a range of entities that the development of indicators can be facilitated as well as being made more manageable. Not all variables originally defined for use in the analyses contribute significantly to all factors. Generally only a comparatively small number of variables have high loadings (ie weightings) on any one factor and thus only these really need be used to define the indicator and to estimate its value. A criterion for choosing which variables to include is very arbitrary. Certainly, however, variables with loadings below about 0.45 should not normally be included if reasonably pure definition is required. If too few variables have values above that, the factor – and therefore the indicator – should be abandoned.

After selecting the variables to define the indicator, three methods of calculating indicator values for each entity can be used. One is to add standardized scores weighted by their factor loadings – in the way outlined above. A second method is to add standardized scores for those variables whose factor loadings exceed the chosen criterion value but not to weight them differentially. This approach assumes that, although weights have been empirically identified, they make very little difference to calculated indicator values and these differences are not worth the extra computational effort involved. For most situations, this is a reasonable assumption even though the approach is not as statistically pure as it should be. The third method is really a derivative of the first. It ignores differences in the mean values for different variables because, as is shown in Section 3.1, these do not affect order when differentiating among entities. The method can be mathematically described as follows:

$$I = \sum_{i=1}^{n} \frac{w_i \, . \, X_i}{S_i} \qquad \textit{(Equation 4.1)}$$

where w_i = weight (or factor loading) for variable i
S_i = standard deviation for variable i
X_i = basic (or raw) value for variable i for a given education system.

In this equation, variables are still weighted by a factor which is constant across all entities for each variable. It is now however the factor loading divided by the standard deviation of the variable which forms the weight. The effect of this change – and thus the difference between this method and the first – is that here the basic or observed value for an education system variable can be substituted into the equation. This makes it much easier to perform the calculation at least in the majority of circumstances. Values estimated for I by Equation 4.1 can have any value depending on what absolute value (X_i) education systems have for any variable.[1] Such a mode of calculation allows perfectly acceptable interpretation of indicator values for planning, monitoring and research purposes. Brandis (1972) is one who uses this method to form indicator values although he forms social class indicators not education system indicators.

The application of this method in a cross-national setting does not really allow the subsequent interpretation of indicator values in a self-referenced way. Norm-referenced interpretation is all that can really be claimed. Changes in average values are not incorporated into the index and these *per se* could affect any calculated indicator value.

Factor analysis is an excellent method both for empirically defining indicators and for identifying variables which should be combined to form valid indicators. It is a very special and powerful technique, however, and it needs to be used with extreme caution and with full cognizance of the problems and the pitfalls involved. A number of researchers have used the technique in an attempt to identify indicators. Adelman and Morris (1967), Rummel (1972), Banks and Gregg (1965) and Russett (1968) are just some of these. McGranahan *et al* (1972), although they attempt to circumvent the actual technique, implicitly

---

[1] The value calculated by Equation 3.4 can be estimated from Equation 4.1 very easily. It is:

$$I \text{ (Equation 3.4)} = I \text{ (Equation 4.1)} - \sum_{i=1}^{n} \frac{w_i \, \bar{X}_i}{S_i}$$

where $\bar{X}_i$ = mean for the ith variable of the n variables defining indicator I.

apply a sort of 'hand held and guided' factor analysis to develop a socio-economic indicator. Part of the discussion in each of Chapters 5 and 6 is based on factor analytically derived dimensions of education systems. In the first of these, the approach adopted is one of using factor analysis results to guide the definition of the kinds of education system indicators which might be useful for planning, monitoring and research purposes. No strict adherence to the absolute results of any factor analysis is either recommended or advocated for all situations.

Other methods

In certain circumstances, empirical methods other than factor analysis can be used to develop indicators from data measured on interval or ratio scales. These include the methods of canonical correlation and multiple discriminant analysis. Both of these techniques are, like factor analysis, specific applications of the general linear model of multiple regression. These additional techniques would not normally be used and hence they are not discussed here.

A note should however be added on one way in which multiple regression analysis can be used to assist in estimating values for indicators already defined. A regression equation is not something different from what has already been discussed — Equation 3.4 is reported in the form of a regression equation where the weights (w_i) are called 'standardized regression coefficients'. Such an equation as this can be used to adjust previously estimated indicator values by taking into account factors which are known to be affecting particular values. This use of regression — one of the few purposes classifiable as aiming towards self-referenced indicators — is found frequently in the adjustment of values for economic indicators. The values for such indicators reported on a regular basis, even yearly, often show a cyclical trend. Unemployment, for example, is known to be particularly high when a school year ends, but it is lower about mid-way through a fiscal year. Reporting gross unemployment figures does not take these known fluctuations into account. A statement might be made by a government that unemployment has risen or fallen in a particular month or quarter. Any such statement is of limited use, however, because the indicator might normally rise or fall in that period. What needs to be identified is the size of the movement which occurred compared to what might be expected by normal seasonal fluctuation. Regression analysis is used to level out absolute fluctuations in such a time series curve. It then provides 'seasonally adjusted estimates' which are very often more useful than the gross data. Naturally there are governments who wish to capitalize on seasonal fluctuations of either decreases or increases. They see a furtherance of their own political purposes very often and they thus refuse to report seasonally adjusted figures. Generally such series are easy to calculate — especially in most Western countries. (For a full discussion on regression analysis, see Spurr and Bonnini [1967] and

Kerlinger and Pedhazur [1974].)

Regression analysis therefore can have a double role in indicator development besides underlying all other multivariate methods. For most empirical derivations of indicators from variables measured on ratio or interval scales, however, the explicit application of factor analysis is by far the most important technique. Regression, in its normal form is most useful for smoothing indicator values which form a time series. Unfortunately in education, very few indicators yet exist to allow time series to be developed.

4.2 Norm-referenced indicator formation from ordinal and nominal scales

The empirical definition of indicators from ordinal and nominal scales has not been undertaken very frequently to date. Partly this is due to the greater attention paid to variables measured on more sophisticated measurement scales. Such variables measure quantities and they thus lend themselves to straightforward statistical manipulation. It is the pre-occupation with variables of this kind which has resulted in many writers most properly accusing education system statisticians of being unconcerned about aspects of quality. Partly also the lack of definition of indicators from variables measured on ordinal and nominal scales is due to the more powerful, more direct and more easily applied methods developed for analysing variables measured on interval and ratio scales. Indeed the availability of these methods frequently involves the redefinition of ordinal scales – particularly to assume they are at least interval in nature. Unfortunately such an assumption can often confound the concept being measured to such an extent that the concept originally defined is transformed to something not really intended or required. The usefulness of any indicator so derived is therefore minimal.

As noted in Section 2.1, education system characteristics tend to be of a global nature rather than to be aggregates. Their values relate to the system considered as a whole rather than to entities or quantities within the system. Two examples of such global characteristics are:

☐ What is the method for certification at the end of second-level? (eg external examination, moderated school assessment or independent school assessment)

☐ How involved is the local community in school activities?

Variables such as these comment on the overall features of an education system. They are of necessity measured on nominal or ordinal scales. Nevertheless, the nature of such variables demands that they should be used where appropriate and not ignored in preference to 'more measurable' variables.

Certainly, on occasions, an ordinal scale can be assumed to possess the properties of additivity of an interval scale. This is particularly so when some Likert scales are used. For example, the classification of the responses for the 'local community involvement' variable might be recorded as:

- ☐ considerably involved
- ☐ moderately involved
- ☐ sometimes involved
- ☐ not at all involved.

The categories could be allocated numbers from 4 = 'considerably involved' to 1 = 'not at all involved'. If the conceptual distance between 'considerably involved' to 'moderately involved' is regarded as being equal to that between 'moderately involved' to 'sometimes involved' and so on, the Likert scale, although ordinal in definition, is ascribed the properties of an interval scale. Indicator formation across a number of such scales for a sample of education systems can then be empirically achieved by methods such as factor analysis. Most questionnaires can be analysed in this way; so too can ratings of system characteristics by groups of experts. Alternatively, a set of ordinal scales can be refined without making the interval scale assumption. This approach involves examining the distributions across divisions within each scale and then deleting or revising items so as to optimize the validity and the reliability of the final set.

There are many situations in which re-definition of an ordinal scale should not or cannot occur. The divisions along the scale nominated above for the method of certification, for example, should not be defined to represent equal conceptual units. The adoption of a moderation scheme for awarding certificates implies a very significant restructuring of educational practice in line with major changes in educational policy and philosophy – certainly much greater than that requiring a modification to move from an existing moderation scheme to an independent assessment approach. Formation of indicators from some variables based on ordinal and nominal scales therefore requires different analytical methods.

One particular method of empirically defining indicators from nominal scales involves ordering the systems being analysed in terms of the kinds of characteristics they possess. Some systems, for example, manifest a number of different kinds of characteristics while others manifest few. Hence it is possible to define an education system indicator on the basis of the relative degree to which one system is differentiated from others. One method of doing this is a Guttman scale analysis. This method involves nominating a particular set of education system characteristics and then ranking systems in order from those possessing most of the nominated characteristics to those possessing least.

Guttman scale analysis

A Guttman scale analysis involves two stages. Firstly it identifies whether the set of nominated system characteristics forms a hierarchy for those systems being analysed. Secondly, it orders the systems into the identified hierarchy. So as to form a hierarchy, therefore, only characteristics possessed by some systems and not others should really be considered unless they are characteristics of fundamental importance.

An example of using this procedure to develop an education system index is provided by Farrell (1969). He attempts to differentiate among systems on the basis of the complexity of their organizational structure. He analyses education systems in both Asia and South America with the method.

Farrell begins his study by nominating a large set of characteristics which education systems may or may not possess. Some of these characteristics are whether or not a system possesses:

- ☐ a ministry of education
- ☐ a university
- ☐ a pre-primary school
- ☐ an economic research institute.

Each system is scored dichotomously on the basis of the possession or non-possession of each nominated characteristic. Farrell arranges his systems in the way represented in Table 4.3. A fairly structured hierarchy of systems is formed by this method for the year 1960 with very few countries showing exceptions to the overall trend. Indonesia, for example, did not have an agricultural research institute where one might have been expected.[1] Similarly Laos did have a pre-primary school which was not an expected institution from a country at that level of education system development. Farrell calculates a coefficient of reproducibility (CR) for his scale as is the normal practice when developing a Guttman scale. This coefficient measures the probability of properly classifying a system in the hierarchy on the basis of its characteristics. The reported value of 0.95 shows the scale is very precise.

Guttman scale analysis is an excellent technique for empirically defining indicators from nominally scaled variables. It has not been used often in education although it has had extensive use in sociology especially for developing indicators of social status. Harper and Richer (1969, 685) note many examples of its use. In a different context Wanderer (1969) develops an index of riot severity for cities using the technique and then relates the indicator values to other variables such as racial composition of the city population, population density and police preparation.

---

[1] Indonesia's agricultural research institute was established in 1962.

	Ministry, etc	Teacher training institute	Inspectorate	Curriculum agency	Secondary vocational school	Secondary industrial school	University	Pedagogical training for secondary teachers	Secondary agricultural school	Pre-primary school	Special education class	Special education school	Research institute, physical-biological sciences	Research institute, social science	Research institute, agriculture	Univ. faculties of chemistry, biology, physics	Research institute, economics	Special secondary industrial school	Univ. faculties, sociology or anthropology
Yemen	x	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o
Nepal	x	x	x	x	o	o	x	o	o	o	o	o	o	o	o	o	o	o	o
Laos	x	x	x	x	x	o	o	o	o	x	o	o	o	o	o	o	o	o	o
Guinea	x	x	x	x	x	x	o	o	o	o	o	o	o	o	o	o	o	o	o
Sudan	x	x	x	x	x	x	x	o	o	x	o	o	x	o	x	o	o	o	o
Tunisia	x	x	x	x	x	x	x	x	o	o	x	x	o	o	o	o	o	o	o
Saudi Arabia	x	x	x	x	x	x	x	x	x	o	o	o	o	o	o	o	o	o	o
Cambodia	x	x	x	x	x	x	x	x	x	o	o	o	o	o	o	o	o	x	o
Burma	x	x	x	x	x	x	x	x	x	o	o	o	o	x	o	o	x	o	o
Iraq	x	x	x	x	x	x	x	x	x	o	o	o	x	o	o	o	o	o	o
Libya	x	x	x	x	x	x	x	x	x	x	o	o	o	o	o	o	o	o	o
Afghanistan	x	x	x	x	x	x	x	x	x	x	o	o	o	o	o	o	o	o	o
Ghana	x	x	x	x	x	x	x	x	x	x	o	o	o	o	x	o	o	o	o
Syria	x	x	x	x	x	x	x	x	x	x	*	*	o	o	o	o	o	o	o
Iran	x	x	x	x	x	x	x	x	x	x	o	o	x	o	o	o	o	o	o
Lebanon	x	x	x	x	x	x	x	o	x	x	o	o	x	o	o	o	o	o	o
Haiti	x	x	x	x	x	x	x	x	o	x	x	x	o	o	x	o	o	o	o
Malaya	x	x	x	x	x	x	x	x	o	x	x	x	o	o	o	o	o	o	o
Morocco	x	x	x	x	x	x	x	x	o	x	x	x	o	o	o	o	o	o	o
Ceylon	x	x	x	x	x	x	x	x	x	o	x	x	o	o	o	o	o	o	o
Thailand	x	x	x	x	x	x	x	x	x	x	x	x	o	o	o	o	o	o	o
Nicaragua	x	x	x	x	x	x	x	o	x	x	x	x	o	o	o	o	o	o	o
Paraguay	x	x	x	x	x	x	x	x	x	x	x	x	o	o	o	o	o	o	o
Liberia	x	x	x	x	x	o	x	x	o	x	x	x	o	o	o	o	o	o	o
Jordan	x	x	x	x	x	x	o	x	x	x	x	x	o	o	o	o	o	o	o
Honduras	x	x	x	x	x	x	x	o	x	x	x	x	o	o	o	o	o	o	o
Costa Rica	x	x	x	x	x	x	x	x	x	x	x	x	o	o	x	o	o	o	o
Ethiopia	x	x	x	x	x	x	x	x	x	x	x	x	x	o	o	o	o	o	o
Egypt	x	x	x	x	x	x	x	x	x	x	x	x	x	o	o	o	o	o	o
El Salvador	x	x	x	x	x	x	x	x	x	x	x	x	x	o	o	o	o	o	o
Guatemala	x	x	x	x	x	x	x	x	x	x	x	x	o	x	o	o	o	o	o
Dominican Republic	x	x	x	x	x	x	x	x	o	x	x	x	x	x	o	o	o	o	o
Panama	x	x	x	x	x	x	x	x	x	x	x	x	o	x	o	o	o	o	o
Bolivia	x	x	x	x	x	x	x	x	x	x	x	x	x	x	o	o	x	o	o
South Korea	x	x	x	x	x	x	x	x	x	x	x	x	x	x	o	o	x	o	o
Philippines	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	o	o	o	o
Peru	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	o	o	o	o
Pakistan	x	x	x	x	x	x	x	x	o	x	x	x	x	x	x	x	o	o	o
Colombia	x	x	x	x	x	x	x	x	x	x	x	x	x	x	o	x	o	o	x
Ecuador	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	o	o
Venezuela	x	x	x	x	x	x	x	x	x	x	o	o	x	x	x	o	x	x	o
Uruguay	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	o
Indonesia	x	x	x	x	x	x	x	x	x	x	x	x	x	x	o	x	x	x	o
India	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
China (Taiwan)	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	o	x
Chile	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	o	x	x
Mexico	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
Brazil	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
Argentina	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x

C.R. = .95; C.R. using only items with 80 per cent in modal category = .92.

* Ambiguous Data

Table 4.3: *Farrell's scalogram of educational structural differentiation: 49 nations, 1960*

Source: Farrell (1969, 309)

4.3 Résumé

This overview of the empirical methods available to define and develop indicators of education systems demonstrates that there are very few methods allowing data to be searched in such a way that indicators are easily identified. Certainly there are many other analysis techniques not included here which identify variables of importance to a particular phenomenon or which search for differences among systems. These other techniques, however, are not relevant here either because they are not multivariate in their approach or because the results they provide need too many inferences to be made to support the existence of a particular indicator. That this is so is not a criticism of the other methods; they are seeking solutions to different problems.

Of the empirical techniques, factor analysis is by far the most important. It is versatile, powerful and, when used properly, it provides good insight into the structure underlying education systems. It therefore must form the foundation of most attempts aiming to define education system indicators by empirical means.

Chapter 5

Indicators Developed for National Education Systems

It is appropriate at this point to establish the present status of education system indicators defined for national systems and the state of their development. Such a review of the available literature allows an evaluation of the extent to which indicator development has been seriously undertaken in connection with education systems planning, monitoring and research. It also enables a number of indicators to be recommended for use with education systems. The review is arranged within the sub-divisions of the theoretical model proposed earlier. This approach makes it easier to determine where most work has been directed. It does have the weakness, however, of making some of the indicators developed before the model was first proposed difficult to classify. When this happens, an indicator is classified according to the perceived intention of the developers.

The attempt to review previous work is plagued by the problem noted by Fanchette in a broader context. He claims that 'there is a great laxity in the use of the term [social indicator] to the extent that most of the social statistics which are usually collected for administrative purposes are being re-baptized "social indicators"' (1974, 8). The same comment applies to education system indicators. In addition to this problem, many researchers and documentalists refer to a variable – sometimes cross-classified by age, sex, region, etc – as an indicator. Fuller (1971), OECD (1973a and 1973b) and Carr-Hill (1973) all adopt this approach. As this is in opposition to the definition of indicators adopted in Chapter 1, listings of 'indicators' in sources such as these are not incorporated into this review. For the same reason, educational data of the type tabulated in minister's reports, development plans, Unesco *Yearbooks*, etc are not referenced directly. Certainly such statistics can be used to indicate the level of very specific features, but they are not the composite indices which are more powerful in describing the level of education system provision in a country.

5.1 Indicators of national education system inputs

Involvement of human resources

Perhaps the first indicator developed to measure education system inputs and which can be properly classified as such is that proposed by Harbison and Myers. It has already been discussed at length in earlier chapters. By combining the adjusted second-level enrolment ratio with the unadjusted third-level ratio, an input indicator measuring the commitment of human resources by a society to these two levels of the education system is formed. Many methodological arguments can and have been discussed in connection with this indicator. Blaug is particularly critical. In reviewing the index, he comments: 'In a sense, it deserves its fame for it embodies virtually every mistake that it is possible to make in international comparisons of income and education' (1972, 68). Despite these methodological weaknesses, the main difficulty with the index is a conceptual one. It restricts attention to students enrolled at the first two of the three levels of the system. It therefore not only ignores first-level commitment – which is an extremely important level in the developing countries especially – but it also concentrates only on students. The other main human resource devoted to an education system ie teachers, is not considered. By omitting both of these concerns, the Harbison and Myers index presents a restricted view of the human resources input by a society into the education system. It also presents a static rather than a dynamic view of those resources.

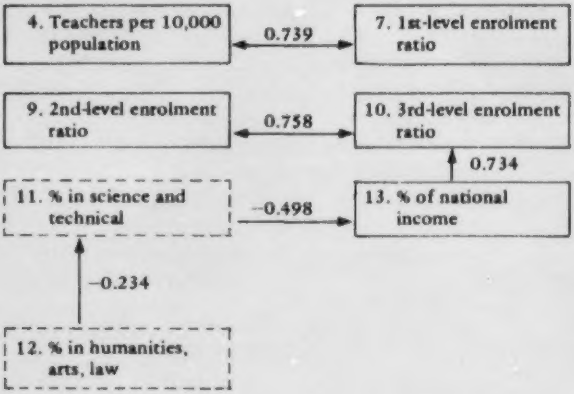
A related input indicator can be derived from the empirical analyses reported in Section 4.1 and in Johnstone (1977). It is termed 'the expansiveness of involvement of a country in educational provision' and it is derived from the first dimension of the factor analyses reported in the two studies. This indicator is very similar to the Harbison and Myers index in that it combines enrolment ratios – or their less error-inducing equivalents – across all three levels of an education system together with a measure of the strength of the teaching force. As its stability has been demonstrated across the 1960 to 1969 period, such a measure must be considered as forming a basis for further development as an indicator. Inter-relating the 1960 factor scores for the indicator (reported in Table 4.2) and the original Harbison and Myers values for their composite index for the same year yields a correlation coefficient of 0.92. This is comparable in value to some of the factor loadings defining the more refined dimension. Hence the Harbison and Myers index, in part validates the use of the 'expansiveness of involvement' concept as a basis for forming an indicator.

The complete Harbison and Myers data can also be re-analysed by applying a simple clustering procedure[1] to the correlation matrix

---

1. The clustering procedure simply identifies those variables whose highest correlations are with each other. Variables whose highest correlation with

reported on page 39 of their study. When this is attempted on all education variables reported, the result shown in Diagram 5.1 is obtained. This shows those variables which are most highly intercorrelated. The teachers and the first-level enrolment ratio variables are thus joined by the very high correlation of 0.739. So too are the second-level and the third-level enrolment ratio variables. A second variable is also associated with the third-level variable – it being the percentage of the national income spent on education. The other two variables shown in the diagram need not be considered at this time because they are classifiable as process variables. This clustering of variables therefore lends support to an argument for the need for an expansiveness indicator.


4. Teachers per 10,000 population
0.739
7. 1st-level enrolment ratio
9. 2nd-level enrolment ratio
0.758
10. 3rd-level enrolment ratio
0.734
11. % in science and technical
−0.498
13. % of national income
−0.234
12. % in humanities, arts, law

Legend

1. Input variables are enclosed in continuous outline; process variables are enclosed in broken outline.
2. An arrow from one variable to another indicates the direction of the highest correlation that variable has with all others. The value of the correlation is written above the arrow.
3. Numbers for variables are those used by Harbison and Myers (1964, 39).

Diagram 5.1: *Summary results of crude clustering on Harbison and Myers correlation matrix*

The results of two other studies which have incorporated a number of education variables can be similarly analysed. These studies are those by Harbison *et al* (1970) and McGranahan *et al* (1972). The former

---

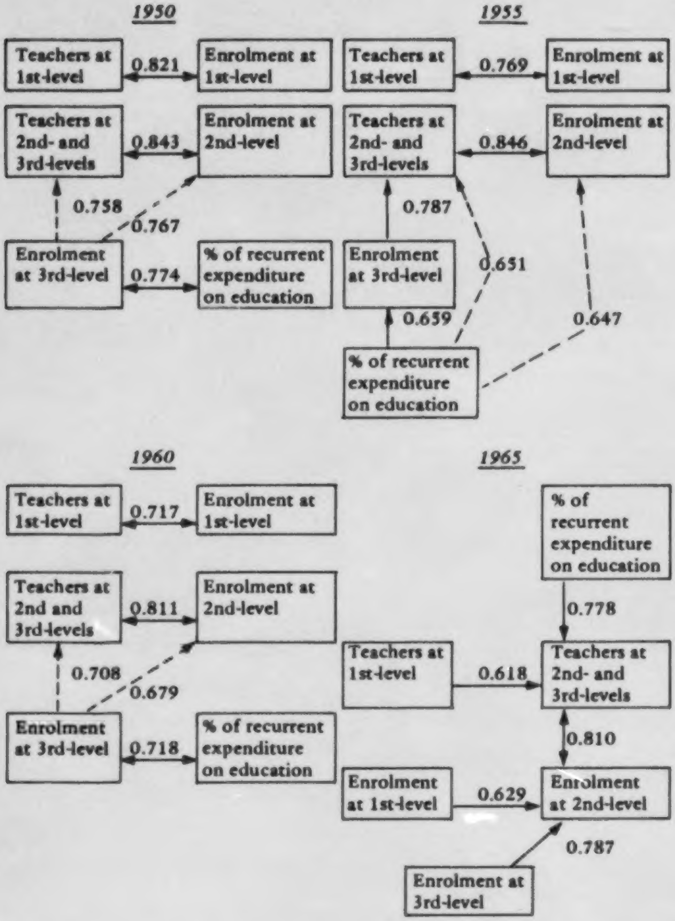
another variable is not reciprocated with a highest correlation are identified secondarily.

reports data for 1950, 1955, 1960 and 1965 across a number of countries, while the latter reports data only for 1960. Summaries identifying those education variables in each study which are most highly inter-correlated are presented in Diagrams 5.2 and 5.3. These summaries are based on the same crude clustering method as was adopted for Diagram 5.1. Those variables which are most highly inter-correlated are those which define the expansiveness of involvement indicator. Hence all of the patterns in the two diagrams confirm the nature of the expansiveness dimension for education system inputs.

Of the two related indicators developed to measure the concept being discussed, the more complex one is to be preferred. It combines more information about the human resources devoted by a society to its education system. It will be referred to from this point as an indicator of the involvement of human resources in the provision of education. The components of the indicator are listed in Table 5.2 and it is recommended, when the indicator values are being calculated, that all components be equally weighted.

Development emphasis on human resource involvement

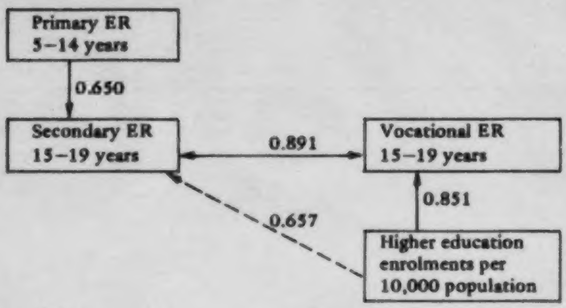
A second input indicator also emerges from the empirical analyses cited above. It measures the development emphasis a country places on increasing its involvement in education and it does this by combining a series of variables measuring the rates at which selected stock variables change in value. In the analysis of 1969 data, it is the rate of development of education at the first level which is emphasized. In the analysis of 1960 data, the emphasis is on the rate of overall development of enrolment ratios. In either of these forms, such an indicator as this is conceptually very different from the involvement indicator described above. Here, flows into the system are being measured whereas before the stock of human resources forms the focus of measurement attention. Although the empirical definitions of the indicator of development emphasis are slightly different in nature in the analysis of data for the two years, a case can be made for creating an overall index covering all three levels of the education system. Indeed the coverage of characteristics defined by the dimensions can be validly extended. As measured by the 1969 dimensions, the emphasis on the rate of change in the involvement of teachers in education is slight. To exclude teachers from this indicator, however, and to concentrate only on pupils is difficult to justify, particularly in view of the first indicator which does incorporate a teacher variable. In order to achieve some similarity between the two potential indicators, therefore, variables measuring characteristics of pupils and teachers should be used. The possibility always exists for disaggregating the indicator into particular levels or into particular features when this is required. The indicator as defined here can thus be termed the 'development emphasis on changing human resource involvement'. Again the components should be equally


1950
Teachers at 1st-level
0.821
Enrolment at 1st-level
Teachers at 2nd- and 3rd-levels
0.843
Enrolment at 2nd-level
0.758
0.767
Enrolment at 3rd-level
0.774
% of recurrent expenditure on education
1955
Teachers at 1st-level
0.769
Enrolment at 1st-level
Teachers at 2nd- and 3rd-levels
0.846
Enrolment at 2nd-level
0.787
Enrolment at 3rd-level
0.651
0.659
0.647
% of recurrent expenditure on education
1960
Teachers at 1st-level
0.717
Enrolment at 1st-level
Teachers at 2nd and 3rd-levels
0.811
Enrolment at 2nd-level
0.708
0.679
Enrolment at 3rd-level
0.718
% of recurrent expenditure on education
1965
% of recurrent expenditure on education
0.778
Teachers at 1st-level
0.618
Teachers at 2nd- and 3rd-levels
0.810
Enrolment at 1st-level
0.629
Enrolment at 2nd-level
0.787
Enrolment at 3rd-level

Note:

1. See Diagram 5.1 for explanation of mode of presentation.
2. Broken arrows indicate relationships almost equal in strength to the strongest.

Diagram 5.2: *Summary results of crude clustering on Harbison et al correlation matrices*



Note:
See Diagrams 5.1 and 5.2 for explanation of mode of presentation.
ER = enrolment ratio

Diagram 5.3: *Summary results of crude clustering on McGranaban et al correlation matrix*

weighted when estimating indicator values.

Financial involvement and expansion

A third input indicator is also identifiable from the 1969 empirically defined dimension set. It measures the 'financial commitment and emphasis' made by a country to a national education system. Two stock variables – the proportions of GNP and the national budget devoted to education – and two flow variables – the rates of change in the values of the stock variables – are combined here. The composite of all four variables is statistically justifiable on the basis of observed correlation coefficients. It is, however, difficult to interpret the final value in planning terms. High values on the dimension, for example, can be calculated because a country:

- ☐ has low expenditure but has increased its relative commitment over recent years
- ☐ has high expenditure and has increased its relative commitment over recent years
- ☐ has very low expenditure and has not changed its commitment very much over recent years.

Such ambiguity in interpretation should be overcome because there does seem to be support from other studies indicating that a financial indicator of some kind is useful for differentiating among countries. Asher and Shively (1969) is one such study. They re-analyse the original Harbison and Myers' data using multiple discriminant analysis and find that one way to distinguish among countries is on the basis of a dimension measuring 'the allocation of economic resources to education' (1969, 185). A re-examination of Diagram 5.1 shows the expenditure variable (13) has an equivocal standing in the pattern drawn. Its direct relationship is with the third-level enrolment ratio, but no conclusion as to its real standing can be inferred here. The possibility thus strongly exists that the financial component is independent of the human resource involvement component. Diagrams 5.2 and 5.3 also reveal a similar pattern of uncertainty concerning the inter-relationship of the financial variable. From this behaviour can be inferred (perhaps *post hoc*) the possible existence of a separate financial input dimension.

Unesco (1974) proposes an input indicator which is an alternative to this empirically defined financial commitment indicator. It measures a country's capacity to finance the development of education. It is defined as follows:

$$I = \frac{f(1 - r)p}{m} \qquad \textit{(Equation 5.1)}$$

where f = first level enrolment ratio
r = public expenditure on education and training as a proportion of total public expenditure
p = public expenditure on education and training as a proportion of GNP
m = total expenditure on education and training as a proportion of GNP.

This indicator combines a measure of the extent to which education in a country is publicly financed (p/m) with the proportion of a national budget not devoted to education (1 − r) and which is thus theoretically available to be diverted to education and with a measure of the proportion of the younger population benefiting from education at the first level. Values for this indicator range from 1.0 (reflecting 'no need for the further quantitative development' [1974, IV/9]) to 0.0 (theoretically reflecting no capacity to finance further development of education). Values less than 0.4 are considered to be indicative of a situation requiring urgent remedial attention.

Although this indicator does have a strong appeal, it is weakened by the ambiguity rendered to interpretation of any value calculated for it. For example, high values approaching 1.0 are calculated in any of the following three situations:

☐ when overall expenditure on education is low but private involvement in education does not exist
☐ when public expenditure on education represents a low proportion of total public expenditure
☐ when the enrolment ratio is high.

Similarly the value can be very low when heavy dependence is placed on private or overseas money for providing education or when the first-level enrolment ratio is low. In addition, educational expenditures benefit more than first-level students. The use of first-level to serve as a weight for the financial constraint component thus excludes this wider group of beneficiaries. With these difficulties, the Unesco indicator as defined cannot be accepted as a viable indicator in any basic input set. It does however support the concept of including a financial input indicator.

Two financial indicators are thus recommended for the input set. One indicator combines two equally weighted variables measuring the stock of funds flowing into education systems – one variable related to GNP and the other related to the total government expenditure. This composite indicator is termed the 'financial involvement indicator'. The second also combines two equally weighted variables – both of the rate of change variables for the two stock variables in the former indicator. This second financial index is thus a 'financial expansion indicator'. Taken conjointly, values for these two indicators comment on the present level of funding for an education system and the emphasis which has been placed on changing that level over the past few years.

Equality of access

Another input indicator proposed by Unesco (1974) measures the difference in the availability and utilization of educational services between rural and urban areas. It is defined as:

$$I = \frac{(a.b)^{\text{rural}}}{(a.b)^{\text{urban}}} \qquad \textit{(Equation 5.2)}$$

where a = first-grade enrolment ratio for new enrolments
b = first-level education completion rate.[1]

The indicator is only proposed for the first-level because 'the distinction rural/urban is hardly meaningful for second-level education, most secondary schools being located usually in urban areas' (1974, IV/4). Unesco points out that it is (theoretically) possible that rural education might be more freely attended than urban education, in which case the value for the indicator as defined by the above equation would exceed 1.0. Although this is highly unlikely, if the occasion did arise, Unesco recommends that the indicator value should be calculated by inverting the value obtained from the Equation 5.2. In either case, values less than 0.5, forming what Unesco terms a 'red zone', indicate a situation needing remedial attention.

An additional indicator analogous in concept to this rural/urban indicator is also defined by Unesco. It, however, measures a very different characteristic of the system. It is concerned with the equality of access to second-level education by pupils from different socio-economic backgrounds. It is defined as:

---

[1] The inclusion of this completion rate variable – measuring a process aspect of a system rather than an input – makes this indicator ambiguous to classify. The intent of the indicator as proposed is, however, an input measure.

$$I = \frac{W^2}{W^1} \qquad \textit{(Equation 5.3)}$$

where W^i = the proportion of pupils enrolled at the ith level whose fathers are farmers.[1]

Unesco chose 'farmers' as a surrogate group for those people in a nation with the lowest annual family income. It is stated, however, that the income measure is preferable if data are available. In many countries, it is probably very difficult to obtain good data for either group.

When the value of the indicator is 1.0, pupils from farmers' families are treated equally in their access to second-level education as they are to their entering first-level education. The minimum value is again 0.0 with values less than 0.5 defined as being indicative of a bad situation.

Indicators measuring the equality of access to education – as in both of the above suggestions by Unesco – can be defined for many kinds of sub-groups. Possible additional indicators could be calculated for different ethnic groups, linguistic groups, handicapped groups, etc. In all such instances, it is desirable to establish an indicator which allows comparability between values. Proceeding in the way Unesco proposes, however, would not allow this comparability. Each indicator of inequality would be defined and calculated slightly differently. The Gini coefficient proposed in Chapter 3 is one way to define an indicator which does allow comparisons among values as well as interpretation against known criterion values. It also enables measures of inequality to be calculated across more than two classes. For example, the Unesco socio-economic inequality of access indicator could be calculated across a number of more clearly defined occupational categories. This would obviate the need to argue for the selection of one type of occupation rather than another.

There are many possibilities which might be considered when defining such an equality of access indicator as is recommended here. Some of these are listed above. In a particular education system, the characteristics of most interest and importance are probably easily identifiable in most cases. Hence the main recommendation made here can only be that in any basic set of input indicators, at least one component should be an indicator measuring the degree of equality of a characteristic – such as urban/rural or socio-economic differences – of importance to educational planners in their efforts to provide an education service to all members of a society. That indicator should be calculated as a Gini coefficient.

One useful indicator of equality of access can be specified explicitly, however. It is defined as the equality of access of girls to education. It can be calculated for either first- or second-level education. Here the

---

[1] Farmers are defined as being workers in ISCO major group 4 although groups 4, 5 and 7 could be used instead if a wider coverage is required.

Gini coefficient compares the characteristic of the percentage of total female enrolment at a given level in each geographical area against the criterion of the percentage of the total population of the appropriate age group who live in each geographical area. An indicator value close to 0.0 demonstrates that each area contributes in accordance with what is expected from it. High values closer to 1.0 demonstrate that some geographical areas are not enrolling as many girls as might be expected. Such a situation would be encountered for example when girls in rural provinces are not encouraged to attend school while girls in the urban areas or the capital city do attend school. This particular indicator therefore measures the relative importance attributed by a society and/ or its political leaders to the education of females. This feature is a particularly important one in many developing countries.

Educational attainment

A final input indicator defined by Unesco (1974) measures changes in the minimum educational attainment. It is:

$$I = \frac{N_t}{N_{t-1}} - z \qquad \textit{(Equation 5.4)}$$

where N_t = the proportion of the total population of an age theoretically corresponding to grade 4 of first-level which is enrolled in grade 4 or above in year t

z = a dummy variable to ensure the value for I is between 0.0 and 1.0.

Minimum educational attainment is therefore regarded as being four years of first-level education – or the equivalent in the non-formal sector. As Unesco notes: 'there seems to be a reasonable consensus that four grades of primary education represents the prerequisite for achieving basic literacy' (1974, III/12). Any value less than 1.0 is defined as representing a critical situation.

Although this indicator purports to measure attainment – which is an output of an education system – it must be classified here as being an input because of the way it has been defined. In essence, Equation 5.4 ratios two specific age/grade enrolment ratios. It does not measure attainment as such. Despite this, the indicator could serve a most important monitoring function in many developing countries. There, for example, plans can be made to increase enrolment ratios to certain levels over a given time period. However, the planned increases can sometimes not be sufficient to counteract the increases in the population aged, for example, between 5 and 11 years – especially if promotion and repetition rates are not changed substantially also. In situations like this, the value of the indicator defined by Equation 5.4 would certainly fall below 1.0. The obvious difficulty in estimating the value is that very often the number in the population corresponding to the selected age is not known. Slight fluctuations in the value calculated for I can thus be attributed to errors of estimation of this population value. Nevertheless the indicator does provide a useful guide to the pattern of inputs into the system and it should therefore be adopted for use. In this adoption, the dummy variable z is omitted so that values of I above 1.0 can be calculated and mapped. Such a mapping would be especially useful for monitoring efforts to increase enrolments in developing countries.

An indicator corresponding to that for change in minimum education attainment can also be developed for second-level education. Instead of minimum educational attainment, however, it would measure change in maximum second-level educational attainment. Its definition is the same as for Equation 5.4 except that now N_t is defined as:

> the proportion of the total population of a country of an age theoretically corresponding to the last grade of the second-level and who are enrolled in that grade or in third-level education in the year t.

The same interpretation is made for this indicator as is made for the previous one. In countries where strands within the second-level are of unequal length, the length of the general education strand is probably the most appropriate guideline to adopt.

This latter index extends the variable which is very commonly used to describe the development of an education system. That variable is the enrolment ratio of 17 year olds (or some other age or age range) as is appropriate. In the developed countries where many more young people are staying on at school longer, such a variable proffers a useful comment on educational involvement patterns. The indicator defined here is not concerned with the level of involvement, however, because such a concept has already been incorporated into the human resource involvement index. Instead, it is concerned with measuring what change has occurred over a given period of time in that specific characteristic. Although both the minimum and maximum educational attainment indicators are very simple indicators in their formation, they nevertheless make important comment on education system functioning and development.

Preferences for education

In the part of the input sub-division dealing with preferences for education, no indicators have really been developed explicitly. Part of the reason for this is that, in attempting such an exercise, the goals some people see as being the prerogative of the education system are really goals for life itself. For example, 'to develop the ability to recognize the rights of individuals' cannot really be classified as being achieved solely within the domain of the education system. It has very much wider connotations. The difficulties discussed earlier regarding the vague nature of many education system goals also make the development of indicators to measure goals held for the education system a much more complex undertaking than is the case for other sub-divisions of the education system space.

Certainly a number of surveys have been conducted in various countries to ascertain the aspirations held by citizens for their education system. These surveys have not been co-ordinated nor have they attempted to work towards a situation which would allow the results to be compared. Consequently it is very difficult to discuss perceived goals for education on a cross-national basis. In addition to this, the nature of the goals narrows the distinction between indicators and variables. For example, a goal may state: 'the system will develop in individuals the knowledge of a foreign language'. For most purposes, it is inappropriate to combine this goal with another stating something quite different. It is necessary to examine people's responses to each statement independently of their responses to other questions. Hence a single variable is being measured, and in such situations it is more appropriate to discuss the single variable than to form an artificial indicator.

A slightly more sophisticated approach to reporting the variable is to measure responses on a 3 to 5 point scale – for example from 'very important' to 'not important' or some appropriate alternative to this. An average response rating could then be calculated and this value used to compare either one goal with another or the assessment of people in one country on a particular goal with the assessment made in another country. Table 5.1 demonstrates this approach with US data from the Seventh Annual Gallop Poll of Public Attitudes Toward Education. Four goals are reported from the survey which asked respondents to rate, on a 3 point scale, the importance they saw for each of a number of goal statements. Calculated mean[1] responses for each goal can enable a better comparison to be made between the goals than is possible by mere inspection of the percentages for each division of the scale. The values in the last column of Table 5.1 confirm the trend established in the percentages in the first column with a slight statistical modification to the first two goals. This confirmation does not always happen, particularly when a scale with more than three divisions is employed. Such a technique as this is fairly common and is used by many researchers for cross-national comparisons. (In a process variable context, for example, see Biddle, 1970.)

Another way to use a set of responses rating the importance, acceptance, etc of a set of goal statements or preferences for education is to measure the change in perceptions of a society across time. Simple comparison of percentages cannot be used here because of the difficulties in equating so many possible movements. Index numbers, especially those incorporating the concepts underlying the Laspeyres index (see Equation 3.12) are particularly useful for this purpose.

---

[1] The rating 'very important' is here given a score of 3, 'fairly important' a score of 2 and 'not important' a score of 1.

Goal area	Percentage of respondents Very important	Fairly important	Not important	Don't know	Mean rating
To know enough arithmetic to figure out such a problem as the total square feet in a room	87	10	2	1	2.83
To have a saleable skill such as typing, auto mechanics, nurse's aide, business machines	85	12	2	1	2.83
To know something about the major nations of the world today, their kind of government and their way of life	49	40	10	1	2.37
To know something about the history of mankind, the great leaders in art, literature, etc	33	44	21	2	2.08

Source: Gallup, (1975, 232)

Table 5.1: *Ratings of perceived importance of four education goal areas – USA*

No indicators within this part of the education system space have yet been properly developed. That this is so is to the detriment of complete indicator usage, especially for planning and monitoring purposes. This lack thus reveals a major area requiring immediate research attention.

Recommendations for input indicators
In light of this review, eight indicators for the input sub-division of the education system space are recommended for planning, monitoring and research use. These are listed in Table 5.2 together with the technical details for their construction. Other indicators must now be developed to measure aspects of education system inputs not covered by this basic set.

5.2 Indicators of national education system processes

Indicators measuring the ways education systems operate – or the ways they process inputs to provide outputs – are much more various than are indicators developed within any other sub-division of the total space. An empirically defined set of process indicators has been proposed (Appendix B and Johnstone, 1977) while a theoretically defined set is

	Indicator name	Variables used in calculation	Conversion of variables	Method of combination
1.	Involvement of human resources in the provision of education	Adjusted enrolment ratios for first- and second-levels and third-level student enrolment per 10,000 population and number of teachers per 10,000 population	1. Index numbers 2. Standard scores	Unweighted average
2.	Development emphasis on changing human resource involvement	Variables measuring the rate of change in each of the variables nominated for input indicator 1	1. Index numbers 2. Standard scores	Unweighted average
3.	Financial involvement indicator	Total amount spent on education as a proportion of the GNP and the total public expenditure on education as a proportion of the total government expenditure	1. Index numbers 2. Standard scores	Unweighted average
4.	Financial expansion indicator	Variables measuring the rate of change in both of the variables listed for input indicator 3	1. Index numbers 2. Standard scores	Unweighted average
5.	Equality of access	Percentage of enrolments with a certain characteristic in nominated groups and percentage of total population in those same groups	None	Gini coefficient
6.	Equality of access of girls to education	As for indicator 5 above but the characteristic is specifically the total enrolment of girls in each region	None	Gini coefficient
7.	Change in the minimum educational attainment	Proportion of the total population of an age theoretically corresponding to grade 4 of first-level which is enrolled in grade 4 or above (for two years)	None	Division
8.	Change in the maximum second-level educational attainment	Proportion of the total population of an age theoretically corresponding to the last grade of the second-level and who are enrolled in that grade or above (for two years)	None	Division

Table 5.2: *Summary of the core indicators for the input sub-division of education systems*

proposed by Unesco (1974). In addition, single indicators have been suggested or used by particular writers.

In the empirically defined set, ten indicators are identified for 1960 and eight are identified for 1969. There is considerable agreement between the sets for the two years, as Appendix B shows. Such demonstrated stability serves to encourage the identification of major dimensions with which to measure educational change. As is discussed in Section 4.1, the two sets of indicators can be used in the format in which they have been defined by the data analysis. Such an approach is demonstrated in Chapter 9. The real contribution of the dimension sets, however, is as an aid in the identification of concerns of significant importance for both planning and research purposes. This kind of contribution is much more important than using the indicators in their factor form. In the next few paragraphs, therefore, indicators are described which stem from the empirically defined dimensions. These are presented rather than a description of the empirical dimensions themselves.

Relative availability of student places

The first process indicator to be discussed and recommended is one measuring the expansiveness of student involvement in education or, described in another way, the relative availability of student places at consecutive levels. The foundation for this indicator emerges very strongly in both of the empirical analyses although not in identically similar forms. Basically this indicator aggregates measures of the relative widths of successive stages or levels of an educational pyramid. It is defined here to take into account the usual three levels of a system as well as pre-school and special education.[1] Hence the variables used are the individual ratios of pre-school, second-level and special education enrolments to the base of first-level enrolments and the ratio of third-level to second-level enrolments. When possible, all ratios should be divided into components separating male from female enrolments. This allows greater accuracy in monitoring system development especially if there are differences between the sexes in enrolment patterns.

The values for all of the ratio-type variables required for this indicator can be expressed as decimals or as percentages. If expressed as a decimal, each value can be interpreted in terms of the number of students at one level per student at the base level; if expressed as a percentage, each value is interpreted as the number of students at one level per 100 students at the base level. For example, a second-level to first-level ratio could be expressed as 0.2531 or as 25.31 per cent. Hence for every 100 students enrolled at the first-level, there are 25.31 students enrolled at the second-level. Many people prefer the 'per 100' interpretation and thus would express the ratios as percentages. Either form still yields the same indicator values, however.

---

[1] In some situations, it may be preferable to exclude pre-school, and/or special education and/or third-level students. Second-level can then be divided into junior and senior stages. Only two variables might then contribute to indicator value:

☐ the ratio of junior secondary to first-level enrolments
☐ the ratio of senior secondary to junior secondary enrolments.

This modification is made in Johnstone (1978a). Any combination of variables which is appropriate for a particular situation and purpose can be defined. Of course all systems being analysed at a given time would need to be measured by exactly the same set of variables.

Values for the index can be determined by combining the component variables by summing the z-scores for the individual variable values pertaining to an individual education system. This approach requires means and standard deviations to be calculable and hence a reasonable number of education systems would need to be examined. An alternative to this approach is to add index numbers together – again after standardization to a pre-determined mean and standard deviation if possible. High values on the indicator reflect comparatively high development at all levels and stages of the system. If one level predominates – as is the case in many of the developing countries – the indicator value would be relatively lower.

A case can be argued that this indicator should be formed by multiplying together the values of all the component variables and obtaining a geometric mean value. An additive model, as defined by Equation 3.1, is recommended, however, because of the extremely low decimal values possible in some countries for some of the variables when the indicator is formed for norm-referenced purposes.

Two associated indicators

Although not empirically derived, two other indicators can be defined to parallel or to act as substitutes for this first indicator. These measure two allied features of education systems and they are based on the same concept of measuring the relative availability of education at consecutive levels. These alternatives are the relative availability of classes at consecutive levels, and the relative availability of schools at consecutive levels. In these two indicators, the component variables are formed by estimating the number of classes or schools, as appropriate, at the pre-school level, second-level and for special education and ratioing each of these to the number at first-level; it also involves ratioing the number at third-level to those at the second-level.[1] Each of these variables is interpreted in an analogous way to the component variables for the previous indicator. That is, each can be regarded as the proportion or percentage of classes (or schools) at one level compared to one or to 100 classes (or schools) respectively at the base level. For example, the ratio of second-level to first-level classes might be calculated to be 0.1133. This implies that for every 100 classes at the first-level, there are 11.33 classes at the second-level. These component variables are combined in the same way as recommended for the indicator measuring the relative availability of student places at consecutive levels.

---

[1] Estimating classes at the third-level is generally extremely difficult and the implications of a count of institutions at the third-level are very different from the implications of a count of the second-level institutions. Hence it is probably better to discard the third-level to second-level ratio from these two indicators.

Both of these additional indicators are used in Chapter 8 to demonstrate how to monitor the development of particular regional education systems. They provide similar trends to that established by the indicator for student places and are highly correlated with it. The alternative formulations are however more amenable to policy recommendation. They are thus more appropriate to use in specifying policies and in measuring developments in the area of implementation. For example, a building expansion program can be monitored to ensure that the school availability indicator does not decrease. If values did decrease, this would point to a situation in which opportunities for education at one level compared to the next lower level might have increased, but opportunities relative to the next higher level have decreased. This may not be serious over a two or three year period but it would be over a longer term.

Indicators are definable to complement each of the above. They measure the rates at which the quantities in each indicator change. The development emphasis on quantitative expansion of the system is thus identified. High values for the indicators reflect relative expansion of one or more levels of a system over recent years. The actual variables combined are listed in Table 5.4.

Female involvement

A seventh indicator, also strongly identified in the empirical analyses, measures female involvement in the provision of education.[1] It combines variables measuring the percentage of female students and teachers at all levels. Empirically this indicator is closely related to the first indicator – at least in the 1969 analysis. The conceptual difference implied by the two indicators however motivates the recommendation to separate them here. It also substantiates the independence of this characteristic to the first which was found in the 1960 empirical analysis. Combining the two indicators (or concepts) to form a single indicator also confuses the very different considerations and makes both interpretation and analysis of education systems development much more difficult. The female involvement indicator is thus recommended for inclusion in the basic set of process indicators. Statistically, values are estimated in the same way as for the previous indicators.

---

[1] This indicator identifies the relative involvement in an education system of females as opposed to males. It could easily be redefined to indicate male rather than female involvement.

Orientation of studies

Two additional process indicators can be defined to measure the concentration of students in particular disciplines or types of study. One indicator is defined for second-level students while the other applies to third-level students. The foundations for both of these are laid in the empirical analyses. The values for each indicator can be calculated by adding transformed percentage values. A better approach is provided by using the C index of concentration.

Indices for either the second- or the third-level may be calculated in the way demonstrated by Table 3.8. This approach requires the percentage of students in each defined strand to be calculated and these values substituted into Equation 3.10. The resulting value for C is then interpretable in terms of whether there is a concentration of students in one strand or whether enrolments are fairly evenly distributed across all strands. No imputation is possible from a high C value, however, as to where a concentration exists. Such a determination would require reference back to the basic percentage distribution.

This approach, involving the use of all strands of education provided at a level, can be modified so as to make the C value reflect the extent of concentration in particular areas. For example, the second-level index can be calculated by estimating only two values – the percentages of all second-level students enrolled in the general education strand and in the teacher training strand – and substituting these values into the equation. A high value (close to 1.0) indicates a strong orientation within a system to one of these forms of education. As systems are generally provided throughout the world, the predominant strand would be the general education strand. Hence, ignoring the vocational and technical education strand in the calculation makes the C index reflect the major orientation of second-level education studies within a system. A low value for C indicates that neither general education nor second-level teacher training is emphasized within the particular education system being examined. The C index as defined here thus accounts for all three strands at the second-level and hence it is more powerful in the comment it makes on the relative concentration than if only one of the percentages was reported as a variable independent of the others.

For the third-level indicator, it is also possible to adopt the approach described above for the second-level indicator and select just two or three disciplines of study for which to calculate the percentages of total enrolment. In many situations, this approach is perfectly acceptable. In other situations, a number of important fields might be eliminated and this would reduce the usefulness of the indicator – especially for norm-referenced comparisons. An alternative approach to estimate a C value, and one that at least partially overcomes the interpretation problem, is to calculate C values for broadly related areas and then ratio these C values. For example, one value might be calculated from the three percentages measuring the proportion of total enrolments in each of humanities, education and fine arts courses while the other might combine the two percentages related to the proportion of total enrolment in each of law and social science courses. Forming a ratio of these two C values does give an indication of whether third-level students within a particular country are more inclined to enrol in humanities/education/fine arts courses or in law/social science courses. Other broad areas such as the natural and physical sciences are still excluded here, although they could be included at the expense of one of the above areas. This would be particularly appropriate if an educational plan specified the development of one broad kind of study rather than another.

To demonstrate the situation being described, the values for the ratio of the C indices calculated for humanities/education/fine arts and for law/social sciences are calculated from the 1971 data reported in Table 3.8. Dividing the law C value by the humanities C value produces the following values:

- ☐ for Cyprus, 0.02 (0.0031 ÷ 0.13)
- ☐ for Japan, 6.62 (0.192 ÷ 0.029)
- ☐ for Laos, 7.03 (0.141 ÷ 0.020)
- ☐ for Saudi Arabia, 1.45 (0.089 ÷ 0.061).

Hence Japan and Laos have a very significant emphasis on law/social science courses compared to humanities/education/fine arts courses, while the reverse is the case in Cyprus. The relative concentration in Saudi Arabia is only slightly in favour of law/social science enrolments.

Attempting to comment on the balance between two broad areas of study such as is done here merely by using the total percentages in each area ignores the relative degrees of concentration within each area and this confounds the interpretation. For example, in Japan, the ratio of total enrolments in the two areas is 1.68 (43.82 ÷ 26.14) while for Laos it is 2.63. Laos therefore has proportionately more enrolments in law/social science per enrolment in humanities/education/fine arts courses than does Japan. However Japan has a greater percentage of its student body in each of the two broad areas. Two ways of looking at the same data therefore lead to different comments. Comparing the concentration measures reveals almost the same degree of concentration for the two countries. The ratio thus counterbalances the overall percentage interpretation against the ratio interpretation. This in turn gives a more stable comment when examining education system development in both norm-referenced and criterion-referenced situations.

The usefulness of an indicator of this kind warrants its recommendation, for inclusion in a basic indicator set. In Table 5.4, one indicator is recommended for the second-level while another is recommended for the third-level. The actual definition of which categories are to be included in the calculation cannot however be stated absolutely. Each situation should be considered with respect to the objectives of the particular measurement operation being undertaken.

Six members of the theoretically defined set of indicators proposed by Unesco (1974) can be classified as measuring processes of a national education system. As with the input indicators discussed earlier, however, a mix of types of variables does occur. Perceived intention is thus again used as the criterion for classification.

Expected number of years of education

The first of the Unesco process indicators measures the rate at which first-level education is provided to a population. It is called a 'first-level completion indicator' and is defined as:

$$I = a.b \qquad \textit{(Equation 5.5)}$$

where a = first-time first-grade enrolment ratio
b = first-level completion ratio.

Unesco considers this indicator makes a double contribution in that it is possible to interpret the value of the indicator as being 'a combined measure of internal efficiency as well as the availability and utilization of educational services' (1974, IV/1). The indicator is only defined for first-level because no implication should be conveyed that every pupil must complete second-level studies.

The confounding of an input variable with a process variable could lead to difficulties in interpretation. The Unesco publication itself points to some of these. The main weakness of the indicator is conceptual however. Although its intent is commendable, the indicator does not comment on what it purports as completely as it might. For example, completion of first-level studies in minimum time implies something quite different from completion in double minimum time. In addition data availability for the first-grade first-time enrolment ratio is limited in many countries. For these reasons, this indicator is not recommended for inclusion in any basic set.

An alternative indicator measuring a concept similar to that inferred by the Unesco index is the expected number of years of education delivered by a system to first-level entrants. The conceptual basis and computing formula for this indicator have been outlined in Section 3.3. When norm-referenced comparisons are to be made and the number of grades of education formally provided is not the same across all systems being examined, the value estimated for the number of years of education expected to be delivered should be transformed. This merely involves dividing the calculated number of years by the number of grades formally provided. Comparisons among systems can then be made directly.

The expected number of years of education to be received by an entrant can be a slightly ambiguous indicator for classification. Harper and Richer (1969, 264), for example, argue for its classification as an output. Defined as it has been here, however, the index comments on the operation of the system. Further, it is formed by combining the process variables of promotion and repetition rates. It must therefore be classified unequivocably as being a process indicator.

The indicator as defined has been used by some researchers although in very different ways. King (1972) uses it mainly by way of example to analyse the education system in England and Wales. He defines 22 'states of life' of which 17 cover the education system. Flows from one state to another are then estimated so that the number of years expected to be spent in each state can be ascertained as well as the various probable paths individuals might follow. Lindgren-Hooker (1974) uses the indicator more deliberately to compare the implications of certain planning proposals on the operation of education systems in some of the Arab countries.

An analogous indicator can also be defined for second-level education. The situation at this higher level is very much more complex in many countries because of the different kinds of education often provided. To calculate an appropriate value, one of two approaches can be used. One approach restricts attention to one kind of education only – say general education – and estimates the values for that kind only. This is quite a proper restriction and it allows valid cross-national comparisons across the most common form of second-level education. It also enables an efficient monitoring of second-level education within a single country. The other approach follows King's (1972) example described above and constructs a very large matrix containing all kinds of education provided. The value calculated from this matrix for the expected number of years of education delivered overall can then be divided by the total number of grades at the second-level. Although this approach involves very much more data and some of the strands might be of very little relative importance, it does offer a complete comment on second-level operation.

Expected number of grades of education

A companion index to the expected number of years of education received is described in Section 3.3 as being the expected number of grades of education completed by an entrant. Whereas the indicator reflecting years measures how long an entrant can be expected to remain in a given level, the indicator reflecting grades measures how far through the level an entrant can be expected to progress. The distinction between these two indicators is very important. It is possible, for example, to spend many years in an education system but not progress through many grades. At the first-level particularly, the different educational experiences offered in the different grades are crucial to the development of literacy and numeracy skills. Merely repeating the same experiences is not the most effective nor efficient way to develop those basic skills. Hence the indicator measuring number of grades complements the indicator measuring the number of years in the comment it makes on an education system. Both indicators should thus be calculated for each of the first- and second- levels of the education system.

Availability of qualified teachers

The second process indicator proposed by Unesco measures the availability of qualified teachers. It is defined as:

$$I = (u_1 (1 - s_1)) \, . \, (u_2 (1 - s_2)) \qquad \text{(Equation 5.6)}$$

where u_i = proportion of teachers with qualifications corresponding to standard requirements or higher for ith level education

s_i = proportion of teaching force at the ith level who are expatriates.

In a country where all teachers are nationals and qualified, the value for I is 1.0. Unesco considers that values for I less than 0.35 are not desirable.

The indicator as proposed is a useful one although it provides a misleading value. Equation 5.6 makes the implicit assumption that the distributions of the qualifications of nationals and of expatriates are identical. This would rarely be the case and, the greater the extent to which expatriate teachers are better qualified, the closer is the value of I to 1.0. If it is of no concern whether the qualified teachers are nationals or not, the indicator could be used, although then the variables s_1 and s_2 are redundant.

An alternative indicator can be developed to measure the Unesco concept. It involves dividing the teacher qualifications spectrum into at least three divisions – unqualified, semi-qualified and qualified. Very often, tabulations are made of the number of local and expatriate teachers – either separately or combined – in more specific categories than just these three. Restricting attention to these, however, allows a C index to be easily calculated. A low value for such a C index reveals an even distribution of teachers across all three categories – a somewhat undesirable situation. A high value reflects a high concentration in one category in particular. In most situations, especially in the developed countries, this concentration would be in the 'qualified' category. The indicator therefore measures the availability of qualified teachers in a country.

Sometimes, when making norm-referenced comparisons among a number of systems at different levels of development, interpretation of the C value is confounded by a concentration occurring in any of the three categories – especially in one of the top two. This can commonly occur especially when examining the concentration of qualifications of second-level teachers. When these circumstances arise, the category creating the confounding can be excluded from the calculation. This will generally be the 'semi-qualified' category. Values for the C index then compare relative concentration in the 'unqualified' and the 'qualified' categories and these are generally more straightforward to interpret.

Distribution of qualified teachers

A complementary indicator to that measuring the availability of qualified teachers is that measuring the distribution of qualified teachers across regions in a nation. This would be based on the Gini coefficient and it would relate the percentage of teachers qualified, cumulated across regions in a country, to the percentage of the total national teaching force contributed by each area. Whereas the 'availability' indicator described above measures the relative concentration of qualified teachers in the teaching force, this distributive indicator measures the extent to which qualified teachers are evenly distributed across all regions in a nation. High indicator values indicate an imbalance in the distribution – possibly because of a tendency for qualified teachers to be located in urban areas. Low indicator values reflect a more equal chance for students in any geographical region of a country to have access to qualified teachers.

Educational flow differentials

A third Unesco indicator measures the difference in educational flow patterns between males and females. It is defined as:

$$I = \frac{(a.b.c.d)^{\text{females}}}{(a.b.c.d)^{\text{males}}} \qquad \text{(Equation 5.7)}$$

where a = first-time first-grade enrolment ratio
b = first-level education completion rate
c = transfer rate from first- to second-level
d = second-level education completion rate.

To ensure a value between 0.0 and 1.0, Unesco suggests that when the value for males is smaller than that for females, the inverse of the value calculated by Equation 5.7 should be used instead. For a cross-national comparison, this only confuses the interpretation. Values less than 0.6 are considered to be not desirable.

As this indicator is an extension of Unesco's 'first-level completion' indicator discussed above, similar weaknesses to those discussed in that context exist here. In addition, the indicator defined by Equation 5.7

examines gross flow only. When the data are available – and they would have to be if the above equation is to be used – it would be much more analytically complete and efficient to calculate the expected number of grades of education completed by males and by females.[1] This could be done as described above and as defined by Equation 3.17. First- and second-level education could be combined if it is deemed to be necessary, but the calculation done separately for males and females.[2] Dividing the female value by that for males then gives an estimate of the equality of the flows through the education system. Values about 1.0 indicate equality of flow; values greater than 1.0 indicate a more rapid flow for females than for males. In this form, the indicator can be included in a basic set.

A fourth indicator proposed by Unesco, of a kind similar to that just discussed, measures regional disparities in educational flows. The variables used are exactly the same. Instead of ratioing females to males, however, this indicator ratios the region with the smallest product of the four variables from Equation 5.7 to the region with the largest. As Unesco notes: 'the disadvantage of this indicator is precisely in the fact that it refers only to two sub-regions (the worst and the best) and leaves aside the situation in other sub-regions which are between the two extremes' (1974, IV/7). Thus if one region forms an extremely high or low value and all other regions are very homogeneous, such a situation is not identified.

Unesco does propose an alternative form of this indicator. It is formed by multiplying the national averages of the four variables together and using this value for the denominator. The numerator is then formed from the average deviations of sub-regions from the national average. This modification does overcome the problem of one extreme region. Indeed it over-compensates and the indicator now tends to diminish the importance of any extreme regions. Unesco finally admits this is not a good solution but concludes 'it may perhaps be envisaged that both the first and the second alternatives be used' (1974, IV/8).

Two solutions to the measurement of regional flows are identifiable. The preferable solution, although the one demanding more of the data, involves the calculation for each region of the number of grades in a level expected to be completed by an entrant. A simple standard deviation statistic calculated on the set of grades expected to be completed then

---

1. Grades completed and not years in the system must be computed here. The Unesco concept involves the comparison between how far through the system males can be expected to progress relative to the corresponding distance for females.
2. If the index is calculated for first-level only, no real problem arises. When second-level is analysed, the difficulty of males and females following different paths can present itself. Calculations could be restricted to just second-level general education to obviate this problem at least partially.

reveals how much variability there is among regions in a country.

The other alternative is to calculate two Gini coefficients – one to measure the equality of the grade 1 enrolments in a level compared to total enrolments and the other to measure the corresponding equality for the final grade in that level. Ratioing the two coefficients gives an indication of the equality of regional flows. This alternative does not demand transition coefficients for each grade but it ignores regional differences at intermediate grades. For cross-national comparisons, the ratio of the two Gini coefficients is to be preferred because of the reduction in the demand for data. For planning and monitoring of single systems, the standard deviation of grades expected to be completed is a very much more appropriate indicator.

Two additional proposals

A fifth indicator proposed by Unesco is alleged to measure the relative first-level educational effort. It is defined as:

$$I = (1 + z) - \frac{h}{b.f.g.k} \qquad \textit{(Equation 5.8)}$$

where h = the average annual first-level teacher's salary as a proportion of per capita GNP
b = first-level education completion ratio
f = net first-level enrolment ratio
g = first-level pupil-teacher ratio
k = proportion of first-level recurrent expenditure spent on teachers' salaries
z = a dummy variable to ensure the value for I is between 0.0 and 1.0.

Unesco claims that, as it is defined, the indicator 'combines a measure of per pupil annual recurrent cost as a proportion of per pupil GNP (h/g.k) with measures of internal efficiency (b) and of availability of educational resources (f)' (1974, IV/12). Values below 0.2 reveal a poor situation.

This indicator is a very mixed quantity. Not only does it combine input and process variables together, but it also confounds three very different components. The units of measurement thus become very confused and interpretation of calculated values is extremely difficult. No re-formulation can be suggested here as stemming from the concept implied by Unesco.

The sixth and final indicator proposed in the Unesco set is a measure of the provision of resources for innovation and reform. It is simply defined as:

$$I = H + L \qquad \textit{(Equation 5.9)}$$

where H = proportion of total educational recurrent expenditure spent on in-service training of teachers, non-formal education and educational research

L = proportion of total first-level educational recurrent expenditure spent on instructional materials and books.

Combining the two proportions based on different amounts yields a totally meaningless quantity. Nevertheless, a value of 0.15 is selected by Unesco as the critical value. Further, data are difficult to obtain on a cross-national basis to allow the addition and, as well, this indicator becomes very close to being a simple variable. Hence, although the concept is again a useful one, this Unesco indicator must be discarded.

Structural differentiation

The study by Farrell (1969) referred to earlier is one which develops and uses a single index to identify differences among national education systems. He develops an 'index of structural differentiation' comprised of such elements as whether a country has a proper ministry of education, a university-level teacher training institution, university faculties in biology, chemistry and physics and a university research institute in education. Farrell arranges a number of such elements into a Guttman scale and finds that he can distinguish among countries very effectively. He claims that, because of this discrimination, 'the level of such differentiation at any one point in time is a very good index of the ability of a system to adapt further' (1969, 295). Farrell subsequently uses his index to classify hierarchically 49 Asian, South American and African countries. He also investigates the relationship between his index and other measures of national development and again finds the index to be of considerable utility in distinguishing among the developing nations.

As data are readily available on both the Farrell index and the empirically defined dimensions on which much of the first part of this section is based, it is possible to determine the inter-relationship between the indicators. Table 5.3 reports rank-order correlation coefficients between the structural differentiation of education systems, as reported by Farrell (1969, 309), and the factor scores as reported in Table B.3 for the empirically defined indicators. There are 20 nations common to both studies. None of the correlations approach a level at which the Farrell index could be regarded as partially defining any of the empirical dimensions. Hence it must be concluded that the structural differentiation aspects of education systems are relatively independent of the other dimensions. Although data for 1960 only are used here, the obtained result is sufficient, especially in the light of Farrell's other work, to add the strucural differentiation indicator to the set of process indicators.

1960 process dimension number	Rank-order correlation
1	0.098
2	0.242
3	0.189
4	−0.041
5	−0.106
6	−0.318
7	0.166
8	0.075
9	−0.268
10	−0.148

Table 5.3: *Rank-order correlations between the structural differentiation indicator and the 1960 empirically defined dimension set*

Dependence on a single source of funds

The usefulness of another process indicator can be posited from one of the indicators recommended by Unesco but one which is considered as an input indicator because of the way Unesco originally defines it. By transforming that original definition (Equation 5.1), an indicator can be defined to measure the dependence an education system has on a single source of funds. Statistically, the indicator is based on the C index of concentration (Equation 3.10). It examines the proportion of total educational expenditure coming from various sources – for example, from national governments, from state or provincial governments, from private sources and from overseas aid. The index does not indicate on which of these (or other) sources a country's education system depends most. It does, however, reflect the concentration or dependence on a single source as is intimated in the Unesco indicator. Hence a 'dependence on a single fund source' indicator can be defined. To ensure some compatibility for cross-national comparisons, two divisions only are nominated for use. One of these sources is the central or federal government of a country while the other is the state or provincial government. The proportions of total expenditure for a national education system coming from both of these sources thus needs to be estimated before the calculation can be done.

This indicator, although measuring sources of funds, is nevertheless still classifiable as a process indicator and not as a system input. The definition in Section 1.5 states that input indicators relate quantities 'taken by an education system to the total amount available for distribution'. This concept in the financial area is thus already covered in the input indicator set by indicator 3 listed in Table 5.2. Here, however, concern is for the total education budget and it is being sub-divided into the various sources. As no relationship is inferred

Indicator name	Variables used in calculation	Conversion	Method of combination
1. Relative availability of student places at consecutive levels	Ratio of enrolments at pre-school to first-level Ratio of enrolments at special education to first-level Ratio of enrolments at second-level to first-level Ratio of enrolments at third-level to second-level (each ratio separately for males and females)	1. Index numbers 2. Standard scores	unweighted average
2. Relative availability of classes at consecutive levels	Ratio of numbers of classes at pre-school to first-level Ratio of numbers of classes in special education to first-level Ratio of numbers of classes at second-level to first-level Ratio of numbers of classes at third-level to second-level	1. Index numbers 2. Standard scores	unweighted average
3. Relative availability of schools at consecutive levels	Similar variables to those above except numbers of schools are used instead of numbers of classes	1. Index numbers 2. Standard scores	unweighted average
4–6. Development emphasis on quantitative expansion of system for students (4), classes (5) and schools (6)	Rates of change for all variables listed for indicators 1–3	1. Index numbers 2. Standard scores	unweighted average
7. Female involvement in the provision of education	Percentages of enrolment who are female at the first-, second- and third-levels Percentages of teaching force who are female at the first-, second- and third-levels	1. Index numbers 2. Standard scores	unweighted average
8. Orientation of second-level studies	Percentages of second-level students enrolled in general education and teacher training courses	No	C index
9. Orientation of third-level studies	Percentages of third-level students enrolled in nominated fields	No	1. C index 2. Ratio of C values

Indicator name	Variables used in calculation	Conversion	Method of combination
10. Expected number of years of education received by an entrant	Promotion and repetition rates for all grades in the first-level and/or second-level	No	Special formula (Equation 3.15)
11. Expected number of grades of education completed by an entrant	Promotion and repetition rates for all grades in the first-level and/or second-level	No	Special formula (Equation 3.17)
12. Availability of qualified teachers	Percentage of teachers in various qualification categories	No	C index
13. Distribution of qualified teachers	Percentages of qualified teachers in each area of a country as well as percentages of total teaching force in each area	No	Gini coefficient
14. Male/female educational flow differential	Promotion and repetition rates for each grade for males and females separately	No	Special formula (Equation 3.17) then division
15. Regional educational flow differential	1. Promotion and repetition rates for each grade in each region 2. First and final grade enrolments and total enrolments for the level analysed	No	1. Special formula (Equation 3.17) + standard deviation 2. Ratio of Gini coefficients
16. Index of structural differentiation	Variety of characteristics of institutions	No	Guttman scale
17. Dependence on a single fund source	Percentages of all expenditure on a system provided by the national government of a country and by state or provincial governments	No	C index
18. Education cost indicator	Unit costs for first-, second- and third-level over a given time period and enrolments at each level in a base year	Index numbers	Laspeyres formula

Table 5.4: *Summary of the core indicators for the process sub-division of education systems*

between each amount and the total amount of its source, the indicator cannot be classified as measuring inputs.

Education costs

An additional process indicator emerges from the review and evaluation presented, although it has not actually been developed in any study to date. It can be called an 'educational cost indicator' and it was discussed briefly in Section 3.3.[1] It is defined as a cost index of a sort similar to that of the commonly used consumer price index. It measures the relative cost of providing education within a country. It is thus not designed to measure the distribution of financial resources across the levels of the education system nor to measure the total costs of producing the educational output. Just as the consumer price index measures the relative cost of buying a 'basket of goods', the proposed education cost index measures how much more the provision of education costs a country independently of considerations related to either changes in the enrolment numbers or changes in enrolment distribution. Where productivity is increasing, this is reflected in a lowering of the ECI value. As elasticities in education are generally greater than 1.0, however, this lowering in value does not occur very often. Any set of calculated values mostly forms an increasing sequence, therefore. The concern of planners is thus how to keep the sequence of ECI values from increasing too rapidly.

Recommendations for process indicators

Summarizing these proposals for the set of process indicators reveals that 18 indicators[2] can be recommended for use in planning, monitoring and researching national education systems. They are listed in Table 5.4 together with the variables on which they are based and the methods for their formation. The indicators can also be used for planning purposes and they would normally then also be supplemented with selected variables to measure particular features of a system. As with the recommended input indicators, the list presented in Table 5.4 is far from complete. It is meant to form a basis to which further developmental work will add additional elements.

Despite this lack of completeness, Table 5.4 demonstrates the claim made in Chapter 1 regarding the efficiency in reporting which can be achieved when indicators are used. Although it is not possible to estimate in general the number of values required for indicators such as numbers 14 and 15, it can be pointed out that the remaining 16 indicators summarize information from over 100 variables. By presenting the indicator values for a given education system in proper ways (see Chapter 7) the implications of all the recommended indicators could be easily comprehended. A valid estimation of the status of that system is then facilitated. Attempts to reach the same conclusion by examining the values for each of the many variables would be almost impossible to make.

---

[1] The education product index described in Section 3.3, although not as useful for monitoring purposes, could replace the cost index when projections of likely expenditures are undertaken in the planning of an education system. Because of their similarity, both of these indices would not normally be used concurrently.

[2] Indicators in Table 5.4 numbered 10–15 can be defined for either first- or second-level. No specification of level is made here, however. By defining all six for both levels thus increases the number of recommended process indicators to 24.

5.3 Indicators of national education system outputs

In the output sub-division of the education system space, there are three broad areas in which indicators might be developed. These are:

☐ the achievement levels reached by those in or leaving the system
☐ the rates at which people leave the system through drop-out or graduation
☐ the satisfaction people have with the education system.

Unfortunately very few proper output indicators have been developed in these three areas. References are instead mainly confined to individual variables measuring such quantities as average achievement test scores and faculty-specific graduation rates. This lack of development is attributed to difficulties in measurement. In this output sub-division of the education system space, there are more intangible quantities (such as levels of achievement) than in any of the other sub-divisions. In addition there has traditionally been a lack of collecting primary data here by means of annual returns. Data are certainly more difficult to define and collect in this sub-division. If attempts to overcome the problems involved do not become more systematic and frequent, however, the evaluation of education systems will continue to be as incomplete as it is now. This is not to imply that measures of output are the most important measures in any such evaluation. They are of equal importance to measures of the input and process aspects of systems. Unfortunately this equality is denied, ignored or contradicted by some writers.

Achievement

The most commonly considered output measure is that of levels of achievement. Despite the attributed supremacy in the importance it has, however, this area was, until very recently, all but lacking in data. During the late 1950s, the International Association for the Evaluation of Educational Achievement project began to examine the possibility of cross-nationally assessing achievement levels. In 1959 a pilot study was undertaken to gather data on mathematics achievement. Over the intervening 18 years, a great deal of research has been conducted. Seven subject areas[1] have now been tested across varying groups of countries, ranging from as few as eight (for French as a foreign language) up to 19 (for science).[2] Twenty-three countries have participated in one or more of the IEA administrations.

Unfortunately the data available to assist in developing indices for education system monitoring are limited. Generally only one administration of each test has been attempted to date and so cross-temporal comparisons cannot be made. In addition, the specificity of the test scores does not allow judgements as to the nature of general educational provision even at the upper secondary level. Indeed the test scores which are produced confuse, to some extent, the distinction between a variable and an indicator. Definitions related to either concept could be used to support a particular argument.

The most use made from the IEA data has been for research purposes. For example, the relationships between school and student characteristics and levels of academic achievement have been investigated in various ways. Some of these analyses could have a direct bearing on planning and policy formation (see, for example, Choppin, 1974). To date, little planning use has, however, been made directly of the results of the test administrations.

One of the principal contributions of the IEA studies is that they have proved it is possible to measure national achievement levels. This has, in turn, provided a catalyst for other efforts in various ways. The NAEP project in the USA – discussed in the next chapter – is one project borne from and encouraged by the success of the first IEA mathematics administration. Various countries have also embarked on national surveys to ascertain the level their students are at in basic literacy and numeracy skills (see, for example, Bullock [1975], Carroll and Chall [1975] and Keeves *et al* [1976]). These efforts will continue and, given time, banks of data will be established to allow indicator development. For this reason, achievement levels in particular subject areas must be defined as being basic components in a core of output indicators. At present, however, the opportunity to experiment with such indicators is extremely limited.

Amount of education

A derivative of this type of indicator does not measure the level of achievement but instead measures the amount or type of education produced by a system. A typical example of such an indicator is that proposed by Ferriss (1969). He defines an academic production index

---

[1] The subject areas are mathematics, science, reading comprehension, literature, French as a foreign language, English as a foreign language and civic education.

[2] Excellent reports on aspects of the IEA project are given by Husén (1967), Postlethwaite (1974) and Passow *et al* (1974).

(API) 'to measure the production of higher education'. A variable reflecting the gross number of graduates produced by a system in a given year cannot be appropriately defined because of the different demands made by different types of degree.

Ferriss identifies five broad kinds of degree and claims that their relative importance – and hence their relative weighting in his index – can be derived from the time spent in completing each degree (adjusted for the first two years of college work): 'time is proportional to the student's effort and to the institutional costs of providing education' (1969, 264-5). The degrees and the weights Ferriss employs are:

☐ a four years bachelors degree – a weight of 1
☐ a five years bachelors degree – a weight of 1.5
☐ a masters degree – a weight of 1.6
☐ an MD degree – a weight of 2.0
☐ a PhD degree – a weight of 3.2.

The index is then defined as:

$$\text{API} = \sum_{i=1}^{5} w_i \,.\, N_i \qquad \textit{(Equation 5.10)}$$

where w_i = weight for each kind of degree as above
N_i = number of graduates in each degree category in any year.

Ferriss claims his API measures 'the cumulative educational experiences completed in a given year' (1969, 265).

In using the index, Ferriss finds the four year baccalaureate degree dominates the value calculated in any given year simply because of the very large numbers of this type of degree which are awarded. On this basis, Ferriss claims that there is a need to redefine the weights used.

The solution to this problem really requires more than just this re-definition. Firstly it requires the values for each variable to be standardized in some way. As the API would be most commonly used for cross-temporal, intra-system comparisons, the best way to do this is to use index numbers. These numbers can then be combined (or averaged) without the numbers of graduates in any one type of degree dominating for that reason alone. Proportional changes can then be compared. Hence, a 20 per cent increase in the number of graduates with one type of degree is equivalent to a 20 per cent increase in another type of graduate regardless of the absolute values actually graduating with each type of degree. The change to index numbers does alter Ferriss' original concept for the API in that the amount of education produced by a system is no longer measured. The index numbers now measure the amount of education produced relative to a base year. This alternative is probably a sounder concept anyway because it is difficult to interpret a value for the amount expressed in absolute terms.

If it is considered that time spent completing a degree should be incorporated into the index value, then weights of the relative magnitude suggested by Ferriss could be used. Alternatively the square or the cube of the weights defined above could be used on the assumption that the higher degrees have greater productivity. Such changes would also reflect the relative difficulty in entering a higher degree compared to entering a first degree. If the square of the above weights was used, a situation closer to that developed by Angell (1951) would be established. This approach, already discussed in Section 3.1, deals with relative frequencies of occurrence in the development of a crime index.

The indicator as defined by Equation 5.10 appears to be a useful planning tool. Ferriss demonstrates this by calculating variations on the basic concept. For example, he computes, for a number of years, the API per 100 graduates. As this quantity increases in value, the number of degrees above the masters level must be increasing. Similarly, the higher education enrolment per API indexes the change in the API which is likely to occur in several years' time. Further, the productivity of higher education faculties can be determined by estimating the API per faculty member. Ferriss calculates a 1949-50 value for the USA of 2.34 as well as a 1963-64 value of 1.51. He concludes therefore that there has been a decline in the overall productivity levels in the USA. Other variations in the definition include student aid per API and higher education expenditure per API. In special planning situations, further functions involving the API indicator could also be defined.

Davis (1964) attempts to develop a more refined concept with which to measure the amount of education produced by a system. He includes a qualitative component by estimating the grade point average of graduates and incorporating this estimate into a defined index. Such an approach is not likely to be successful, however, because of the very wide variety of standards among colleges and universities. Warren (1971) and Pinchak (1974) are two among many studies to demonstrate the extraordinarily wide variety of grading practices operating in higher education institutions.

The approach adopted by Ferriss and based on the number of years of education received by those people successfully leaving an education system seems to offer most promise for an indicator measuring the amount of education produced by an education system. The concept is an important one and it need not only be applied to third-level education. At the second-level, where there are often alternative types of education offered, an API could be calculated for the whole level or separately for just the junior or senior levels. Weights based on the length of formal training would not be really useful because of the similarity in lengths for most if not all types of education on offer. Instead, some weighting based on stated policy objectives should be used.

Specialization of graduates

The field of specialization of graduates produced by a system is not a quantity which can be properly measured by an indicator. This is more properly the task for a variable especially if the achievement of planning targets is to be evaluated or monitored. What indicators can measure, however, is the mix of graduates across fields of study. This then is the same kind of measure as is defined for the process indicator measuring the orientation of third-level studies. Indeed the use of a C index here in the output sub-division reveals if the relative mix of graduates is similar to that for the student body. In some cases, differences will certainly be detectable and explainable – a new course may have been recently introduced and no graduates may have yet been produced. In other situations, a difference between the process and output C values would point to different flows of students through the respective faculties. In general, the more a system develops, the lower should be the value of the C index. A high value for C indicates a specialization of the graduates in a certain area – something which does not happen when a variety of faculties are each providing a number of manpower skills to the work-force of a country. The adoption of an indicator measuring the degree of concentration of specialization of third-level graduates is therefore recommended. A parallel indicator for the second-level is difficult to justify, however.

Students leaving the system

The second area of the output sub-division in which indicators might be developed is that of the rates at which students leave the system. Considered in isolation, such rates are really variables for they apply to each level independently (or even to each grade if considered in the extreme case). When viewed in this way, drop-out and graduation rates comment too specifically on output as defined. A more complete view of output is provided if the number of students graduating from each level is related to the number who entered that level. In terms of the original definiton of output, therefore, the 'similar characteristic' should be entrants while the output measurement is 'ever graduated'. The output indicator can thus be termed 'the proportion of entrants who finally graduate'. It was defined by Equation 3.19 and described in Section 3.3.[1]

Although many studies cite the variable of graduation rate as making an important comment on the output from an education system, no major published study uses the composite indicator. Nevertheless, it should be included in a basic set for planning and monitoring education system development. Values for the indicator should be estimated separately for first- and second-levels and, where possible, separately for males and females. The nature of third-level education diminishes the applicability of this indicator in that context except in very special circumstances.

---

[1] Criticism could be made of this indicator because it combines process and output variables. This is valid criticism especially in view of some of the comments made earlier. The process variables, however, only operate to reduce the number who enter to the number who (theoretically) enter the final grade. Hence the indicator still conforms to the definition of output originally stated. In addition, the concept imbedded in the indicator is widely accepted as being an output.

Satisfaction with the system

The final area of the output sub-division in which indicators might be defined concerns the satisfaction of society with an education system. Indicators here measure the effectiveness of the system as seen by various groups of people – for example by students, parents, employers, etc. To measure this satisfaction properly, it would be necessary to construct a scale – perhaps with Likert-type responses – containing a number of items. Each item would then measure the respondent's satisfaction with specific aspects of a particular system. For example, one item might measure satisfaction with the development and maintenance of literacy skills; another item might measure satisfaction with the development of individual self-sufficiency, another with the ability to think critically and so on. A respondent's answer to these questions could then be combined into an indicator measuring overall satisfaction. Attitudes to particular features could of course still be analysed in the way component variables of any indicator may be measured.

Unfortunately no real attempt has been made to measure the satisfaction with education systems in a coherent way. The annual Gallup polls of Public Attitudes toward Education are perhaps the most ambitious attempts in this direction although they are really measures of variables and not of indicators in the sense implied here. In the Gallup polls, the following question is asked of the respondents in the sample: 'Students are often given the grades A, B, C, D and FAIL to denote the quality of their work. Suppose the *public* schools themselves, in this community, were graded in the same way. What grade would you give the public schools here – A, B, C, D or FAIL?' (Gallup, 1976, 188) . Table 5.5 reports the results of the ratings obtained in 1974, 1975 and 1976 cross-classified by whether respondents had children in school and, if so, what type of school the children attended.

There has been considerable movement over the past few years in the percentage of respondents within each category. Generally the movement is towards a rating of moderate preference for the US public school system. The numbers of people awarding an A grade have dropped substantially in each of the three special categories, although so too have the proportions of respondents awarding the D and FAIL grades. The mean ratings in the last row of the table point to an overall disillusionment of respondents with the system – the average rating is

	No children in school Year			Public school parents Year			Parochial school parents Year			National total Year		
	74	75	76	74	75	76	74	75	76	74	75	76
A	16	11	12	22	17	16	12	5	6	18	13	13
B	22	26	26	22	36	34	28	25	28	30	30	29
C	19	27	26	42	29	30	22	24	40	21	28	28
D	7	9	9	24	8	10	9	16	12	6	9	10
FAIL	4	7	7	4	7	5	15	18	4	5	7	6
Don't know/ No answer	32	20	20	5	3	5	14	12	10	20	13	14
Mean	2.57	2.31	2.34	2.76	2.49	2.48	2.15	1.81	2.22	2.63	2.38	2.38

Source: Gallup (1974-76)

Table 5.5: *Gallup poll ratings of US public schools 1974-76*

just above the grade of C. The frequently quoted teacher's comment of a pupil 'has potential but there is considerable room for improvement' thus seems to apply to the US public school system as a comment from the public it serves.

Data like these provide very valuable information for policy-makers and planners. Further inquiries must of course be made to determine why the dissatisfaction exists. On the basis of these data, action can be taken in specific areas and the perceived effect of the implementation of the ensuing policies can be monitored over the following years. It is for this reason that an indicator measuring the mean rating of an education system by specific groups of respondents must be defined and incorporated into a basic set of output indicators. Admittedly there are many difficulties with such an indicator but, in the absence of any better approach, the mean rating from a Gallup-type poll survey must be included.

One further observation can be made from the values in Table 5.5. It concerns the very large numbers of respondents who answer the satisfaction question with 'Don't know' or 'No answer'. Indeed, except for the group of 'public school parents', the percentages in these categories exceed the percentages of respondents awarding an A rating. These large percentages reflect the fact that many people do not consider they can offer proper comment on matters of this type. If questions become too technical or specific, the percentage of such people in a sample rises and the amount of information available to be analysed is reduced. However, an administrator or a planner often wishes to determine reactions to specific questions. A delicate balance is therefore called for between the amount of useful information to be gathered and specificity of content. Perhaps with increased use of surveys to gather data, the public at large may become more discriminating in their views and feel more confident to express their opinions on particular matters of importance.

From the data in Table 5.5, it can also be argued that 'parochial school parents' form a biassed sub-group (in the sampling sense) and that many respondents classified as having no children at school would not be in a position to judge the public schools' performance. These allegations may be quite valid for many such people. They point to the need to identify characteristics of respondents because, in differing circumstances, comments from one particular group might be considered more seriously by planners and administrators than comments from another group. Multiple classifications are also useful at times. For example, do people who are employers in the three sub-categories of Table 5.5 differ in the grades they award from those who are employees or not employed? Such questions tap different kinds of consumers of the products of the education system.

Another way of measuring satisfaction with an education system was described in Section 3.3. It is based on variables measured on nominal scales rather than on ordinal scales such as is the case with the data in Table 5.5. In Table 3.11, for example, the eight major problems perceived to confront US public schools over the 1970-76 period are reported. Relative change in the order of the importance placed on these variables by respondents can be measured by the Laspeyres index.[1] No example of this approach can be cited here but it is an approach which offers considerable possibility for monitoring changes in perceptions to specific problems.

Recommendations for output indicators

The proposals considered here within the output sub-division of the education system space suggest that five indicators can be defined to measure performance across the three broad areas noted earlier in this section. These five indicators are listed in Table 5.6. The defining of only five indicators is not, in itself, cause for concern. These indicators map most of the main considerations towards which educational planners are oriented. A point which should be of concern, however, is that the data on which the indicators are based are available in very few countries. Proper evaluation and monitoring of national education systems are severely impeded, therefore. Efforts must be made in many countries – developed and developing alike – to rectify this serious deficiency. If such efforts are not made, complete educational planning efforts will retain their present shackles for a long time into the future.

---

1. In the example in Section 3.3, initial ranks are used as the weights. An alternative approach would be to use weights derived from official statements of education system objectives.

	Indicator name	Variables used in calculation	Conversion	Method of combination
1.	Achievement level	Scores on either individual items or on sub-tests	No	Addition
2.	Academic production index	Numbers of third-level graduates obtaining various types of degrees *or* Numbers of second-level graduates completing various types of education	1. Index numbers 2. Standard scores	Weighted or unweighted average
3.	Specialization of graduates	Proportion of third-level graduates in nominated fields of study	No	1. C index 2. Ratio of C indices
4.	Proportion of entrants who finally graduate	Promotion, repetition and graduation rates for all grades in the first- and/or second-level	No	Special formula (Equation 3.19)
5.	Satisfaction with the performance of education systems	Ratings on an ordinal scale to indicate degrees of satisfaction *or* responses on a nominal scale to specific features, problems, etc of the education system	No	Weighted average *or* Laspeyres formula

Table 5.6: *Summary of the core indicators for the output sub-division of education systems*

5.4 Indicators of education system-societal congruence

In this final sub-division of the education system space, very many different aspects of society can be considered. The most important of these, from the point of view of the direction in which education outputs flow, are the labour force, societal behaviours and the economy. Certainly there are many more – for example the health sector, politics, family life and planning and the awareness of and the involvement in the community itself. A number of attempts have been made to estimate the relationship between the quality and quantity of educational outputs and the development of all of these aspects. Much of this work is not very conclusive. Hence, little basis is established for empirical indicator development. Certainly a large number of variables can be defined to measure some of these aspects and the list developed by OECD (1971) covers many of these. No attempt has been made, however, to reduce these many variables to more manageable composite indicators.

Educational profile of the unemployed

Two indicators of the congruence between the education system and the labour force are proposed by Unesco (1974). The first of these measures the educational profile of the unemployed. It is defined as:

$$I = \frac{R_i}{R_{\bar{X}}} - z \qquad \textit{(Equation 5.11)}$$

where R_i = unemployment rate of the population aged 15–64 with education level i

$R_{\bar{X}}$ = national average unemployment rate of the population aged 15–64.

The dummy variable z ensures the value for I lies between 0.0 and 1.0 by reducing all values for I above 1.0 to that maximum value. The calculated value is 1.0 when the unemployment rate for those with education level i is the same or higher than the national average. Unesco recommends that the indicator be calculated for the group consisting of those people with no schooling. Any value less than 1.0 then signifies a situation needing immediate attention because it points to the existence of a pool of educated unemployed.

In its present form, the indicator is fairly limited in scope as well as being insensitive to differential unemployment rates at various education levels. The same concept can, however, be measured by a Gini coefficient. When this is the indicator formed, the above difficulties are overcome. To calculate the Gini coefficient in this situation, the cumulative percentage of people in the population aged 15–64 in each education category (from the highest to the lowest) is plotted as a Lorenz curve against the cumulative percentage of total unemployed contributed by each category. The Gini coefficient therefore measures the degree of equality of unemployment across all levels of education. As Unesco warns, the indicator 'should only be used in countries where unemployment statistics are fully reliable . . . there is often a danger that those with no education are less fully covered by unemployment surveys than other categories' (1974, IV/13). In its amended form, this indicator of the educational profile of the unemployed can be accepted into a basic set of indicators for this part of the education system space. When the value is calculated as described above, then usually the higher the value for the Gini coefficient, the greater is the relative rate of unemployment among those with no education or only first-level education compared to those with higher levels of education.

Teacher supply

A second indicator proposed by Unesco for the education system-societal congruence area measures the supply of and demand for primary school teachers. It is defined as:

$$I = \frac{S.T}{z} \qquad \text{(Equation 5.12)}$$

where S = current vacancies as a proportion of those leaving the system with qualifications to teach in primary schools

T = proportion of those having successfully completed, in the previous year, their training as primary school teachers and who joined their profession

z is a dummy variable corresponding to the following conditions:

$z = 1$ when $S \leqslant 1$ and $z = S^2$ when $S > 1$.

The values of I again range from 0.0 to 1.0 although, as with so many of the Unesco indicator proposals, the two different computing formulae give values which are very difficult to interpret unambiguously. Nevertheless, Unesco (1974, IV/14) recommends a 'red zone' as being formed by values less than 0.5 while values equal to 1.0 indicate an optimal situation. It is also pointed out that, if the data are available on both the component variables, the indicator can be redefined for any other occupation/education category.

There are two difficulties with using this indicator. One is the difficulty of interpretation noted above. This is a major problem and must immediately limit the possible usefulness of the index. The other difficulty is that of identifying what a current vacancy really is. In some countries, estimates of demand are made by using a simple formula – often a single variable such as a desirable pupil-teacher ratio. If an insufficient supply of teachers is located, the value for that variable is simply raised a little and a new demand value estimated. In other countries, either permanent teachers teach classes in two or more schools (a very common practice in developing countries) or a large pool of casual teachers is available to fill the vacancies. In this last situation, such people may not wish to accept full-time employment if the option of casual teaching is available. Hence an apparent vacancy in a statistical tabulation might not be a real vacancy in slightly different employment circumstances. For both of these reasons, this Unesco indicator cannot be included in a core set.

Availability of the mass media

New indicators have been proposed to measure the congruence between the education system output and the behaviours exhibited by members of society. In contrast, very many individual variables are frequently cited. One indicator, which is fairly typical of those which have been proposed is that from Unesco (1974) measuring the availability of the mass media. It is defined as follows:

$$I = \frac{A + B + C + D}{500} - z \qquad \text{(Equation 5.13)}$$

where A = number of radio receivers per 1000 population

B = number of television receivers per 1000 population
C = circulation of daily newspapers per 1000 population
D = circulation of non-daily newspapers per 1000 population
z is a dummy variable.

The value in the denominator is fairly arbitrary and is calculated by 'taking two-thirds of the sum of the highest values on each of the four elements reported in any one of the developing countries and rounding the result to the nearest hundred . . . it was felt simply that two-thirds of the sum of the highest values could well be considered as something like a "norm" for a "satisfactory" situation in the availability of mass media' (1974, IV/3). Values less than 0.1 are considered to be in the 'red zone'.

This indicator measures the impact education has on the behaviour of the people in a society. As the level of education in a population rises, the more freely available and demanded are radios, televisions and newspapers. Other variables could also be included as components in this indicator. The number of public libraries per 1000 population is one such variable; the number of museums and art galleries per 1000 population is another.

As a measure of the consumption by a population of the mass media, the Unesco indicator can be accepted into a common core of indicators. Its definition needs to be altered, however. The dummy variable is not necessary. Also the addition of raw variable values is not appropriate for, as is demonstrated in Section 3.2, variations in the standard deviation of these variables will cause one variable to be weighted more heavily than another. As there is no reason to suspect any major difference in the importance of these variables,[1] each should be equally weighted. Calculation of the indicator value can thus be effected in two ways:

- ☐ use scores standardized for each variable across a group of societies or across all societies being considered
- ☐ calculate index numbers for each variable for a given society over a series of years.

The values calculated for each variable in either of these ways can then be summed (or multiplied) and an average formed. The resulting composite value is easily interpreted in either a norm-referenced or a criterion-referenced way.

Rates of return

The third of the major areas nominated as being most appropriate for measuring the congruence between the education system and the society is the economy. As no significant financial output comes from most education systems, the direct impact of any education system on the economy must be measured in terms of equivalent investment. The most appropriate indicator to use here is the rate of return on the investment made in education. This rate is derived from an analysis of costs incurred and benefits derived from education. It therefore cannot be regarded as being a single variable. It is definitely a composite indicator for it summarizes a considerable amount of information and reports this summary as a single value.

---

[1] It is desirable, in any case, to be able to define the weight any variable would have, either theoretically or by guided empiricism.

The rate of return approach attempts to measure the net relative financial benefits derived by obtaining educational qualifications to a certain level. It offsets increases in earnings attributable to the higher qualifications against the investment made in obtaining them. Rates can be calculated either for the society which provides the education system or for the individual who undertakes the education course. The type of outlays made by a society or by an individual and their relativities are quite different. Hence different values are generally calculated for social rates of return as opposed to individual rates of return.

As many estimates need to be made so that a rate can be calculated, the estimated values can sometimes be criticized quite severely. One of the major estimates which needs to be made, for example, is how much extra income over a person's working life can be attributed to higher qualifications. When data are available to guide this estimate (and this is infrequently), they refer to past income differentials and not to future possibilities. It is also known that future patterns must be different because of different occupation/education balances and changed capital investment patterns. Nevertheless, rates of return do provide an excellent indicator of the relative worth of investing in education or in an alternative kind of investment.

Rate of return analyses have been undertaken and/or discussed by many researchers, eg Blaug (1965), and Thias and Carnay (1969). Psacharopoulos has published a series of studies looking particularly at the cross-national implications. In his 1972 study, he summarizes the results of a number of previous studies performed in individual countries. Table 5.7 shows this summary. Social rates of return are generally lower than the individual rates and both are generally higher for first-level than for higher education.

Considerable work has been done to improve the theoretical background of rate of return analysis. It now has, therefore, a comparatively sound foundation. Hence rates of return for each level of education must be included in any core set of indicators for this part of the education system space.

Recommendations for education system-societal congruence indicators

From this overview, five indicators can be recommended as forming a basis for measuring the congruence between an education system and the society into which its outputs flow. These are summarized in

Country	Year	Social			Private		
		Primary	Secondary	Higher	Primary	Secondary	Higher
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
United States	1959	17.8	14.0	9.7	155.1	19.5	13.6
Canada	1961	–	11.7	14.0	–	16.3	19.7
Puerto Rico	1960	20.9	23.8	16.0	*	24.4	23.0
Mexico	1963	25.0	17.0	23.0	32.0	23.0	29.0
Venezuela	1957	82.0	17.0	23.0	–	29.0	27.0
Colombia	1966	40.0	24.0	8.0	†	32.0	15.5
Chile	1959	24.0	16.9	12.2	–	–	–
Brazil	1962	10.7	17.2	14.5	11.3	21.4	38.1
Great Britain	1966	–	3.6	8.2	–	6.2	12.0
Norway	1966	–	7.2	5.4	–	7.1	4.5
Sweden	1967	–	10.5	9.2	–	–	10.3
Denmark	1964	–	–	7.8	–	–	10.0
Netherlands	1965	–	6.8	6.0	–	8.4	10.4
Belgium	1967	–	–	8.6	–	–	25.0
Germany	1964	–	–	–	–	–	4.9
Greece	1964	–	3.0	8.0	–	5.0	14.0
Israel	1958	16.5	6.9	6.6	27.0	6.9	8.0
India	1960	20.2	16.7	12.7	24.7	19.2	14.3
Malaysia	1967	9.3	12.3	10.7	–	–	–
Philippines	1964	8.0	21.0	11.0	8.5	28.0	12.5
Japan	1961	–	5.0	6.0	–	6.0	9.0
South Korea	1967	12.0	9.0	5.0	–	–	–
Thailand	1970	18.5	11.0	11.0	26.5	13.0	14.0
Hawaii	1959	24.1	4.4	9.2	*	5.1	11.0
West Nigeria	1966	23.0	12.8	17.0	30.0	14.0	34.0
Ghana	1967	18.0	13.0	16.5	24.5	17.0	37.0
Kenya	1968	21.7	22.9	8.8	32.7	35.2	27.4
Uganda	1966	66.0	28.6	12.0	–	–	–
N. Rhodesia	1960	12.4	–	–	–	–	–
New Zealand	1966	–	19.4	13.2	–	20.0	14.7

* over 100.0 † over 50.0 – Not available

Source: Psacharopoulos (1972, 56)

Table 5.7: *Social and private rates of return by educational level*

Table 5.8. Certainly there are many more possible indicators which could be defined theoretically, and hence the indicators listed in Table 5.8 form something of a biassed narrow set. It is considered, however, that more empirical work needs to be undertaken before more indicators can be classified here. Because this is such an important part of the total framework of education systems, it can only be hoped that this further work is undertaken soon and that some sound indicators are forthcoming.

Indicator name	Variables used in calculation	Conversion	Method of combination
1. Educational profile of the unemployed	Percentage unemployment rates for selected educational attainment groups and percentage of 15–64 age group in each of these groups	No	Gini coefficient
2. Availability of the mass media	Rates per 1000 population for the availability of radios, television, daily newspapers and non-daily newspapers	1. Standard scores 2. Index numbers	1. Addition 2. Multiplication
3. Rates of return for first-level education	Many estimates including income foregone, forecast earnings, etc. (See Psacharopoulos (1975), Thias and Carnoy (1969), for full details.)	Various	–
4. Rates of return for second-level education			
5. Rates of return for third-level education			

Table 5.8: *Summary of the core indicators for the education system-societal congruence sub-division of education systems*

5.5 Technical difficulties in computing indicator values

When computing values for the indicators defined in this chapter as comprising the basic sets for planning, monitoring and researching education systems aggregated to a national level, a number of general problems may sometimes arise. It is important therefore to note certain guidelines for handling these problems before leaving the present discussion. A common approach to solving these problems should be established for the development of indicator sets across various education systems if comparability of values is to be effected.

Calculation of rates of change

The largest problem arises in the estimation of values for the 'rate of change' variables included in a number of the indicators. The concept behind these variables is that they should measure the annual change which has occurred in the values for particular variables. To estimate any of these values it is not appropriate simply to measure change occurring from one calendar year to the next. Many fluctuations can occur in such a short period for reasons other than those governed by significant policies. Accounting data being recorded in one year rather than another and sicknesses on a day designated for a census of schools leading to a gap between attendance and enrolment are two such reasons. Planned expansion in comparatively large steps effected every two or more years apart would also cause unreliable estimates of yearly change to be made. Normally, therefore, values for the 'rate of change' variables should be calculated from a basis spanning a number of years. Two specific questions must therefore be answered:

☐ how many years should form the basis for the calculation?
☐ what method of calculating average change values should be used?

If too many years are considered for the calculation period, the reverse of the problem noted for the basis of the single year arises. Here the effect of policies over too long a term are being measured. Short term increases or decreases can thus be dampened out of the estimates of the change which has recently occurred in a system. It is recommended that four years provide a reasonable period of time to allow change to be measured. It is a commonly used value for short term plans and does facilitate very easy computation in most of the available estimation methods. Hence the 'rate of change' in a particular variable for 1975 would be estimated over the 1971 to 1975 period. The 1971 value thus changes four times until the 1975 value is recorded.

In some data tabulations for some countries, values at the beginning and the end of such a four year time span are not available. In such circumstances, an appropriate alternative rate, spanning either a smaller or a larger number of years, must be used. In order to obtain maximum comparability between these values, therefore, changes must be calculated as average annual changes and not as percentage gross changes.

Estimates of the average annual change can be made in at least four main ways. To discuss these, Table 5.9 reports some hypothetical data for a variable while Diagram 5.4 shows this data and some of the alternative calculating implications graphically. The possible errors introduced by calculating rates based only on consecutive years is demonstrated here. Major fluctuations both up and down are noted over the sequential years. Considered over the period of time represented there, however, a gradual but significant increase is shown to occur.

One method of estimating the rate of change is to ratio the final value for a variable to the initial value and divide the resultant change by the number of intervening years. That is:

$$\mathrm{ROC} = \frac{\frac{F}{I} - 1}{N-1} \qquad \textit{(Equation 5.14)}$$

where ROC = rate of change in the variable
F = values for the variable in the final year
I = value for the variable in the initial year
N = number of years for which data are being considered.

Year number	Observed variable value	% annual change	Estimated variable values provided by			
			overall mean	geometric mean	annual mean	regression line
1	50	–	50	50	50	49.4
2	54	+8.00	53.3	53.0	53.1	51.9
3	52	−3.70	56.5	56.1	56.5	54.4
4	53	+1.92	59.8	59.5	60.0	56.9
5	63	+18.87	63.0	63.0	63.8	59.4
Estimated rate of change over period			6.50	5.95	6.27	5.00

Table 5.9: *Example demonstrating the implications of various methods to estimate rates of change*

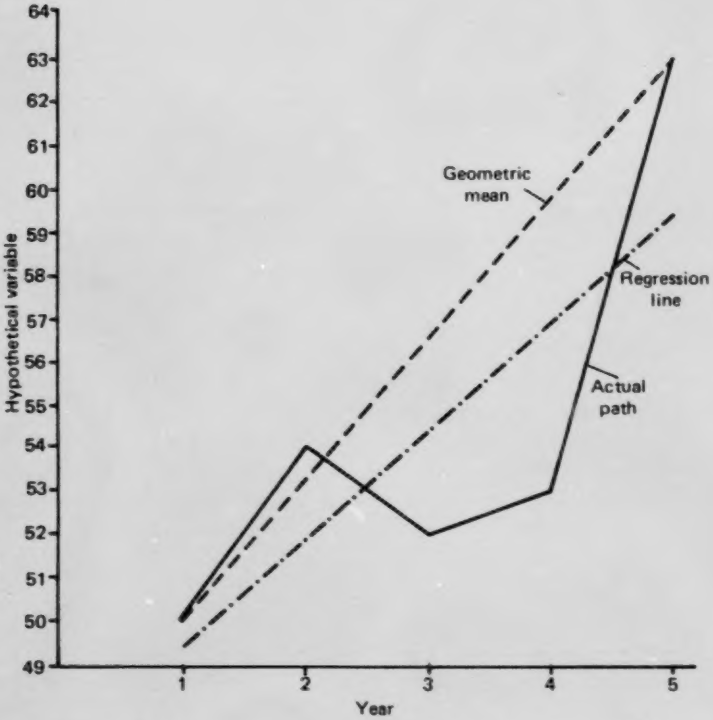


Diagram 5.4: *Comparison between methods for estimating rates of change*

For the data in Table 5.9, this method estimates an average rate of change in the value of the variable of 6.5 per cent. If the value for either the initial or final year is very different from the values for the years on either side of these focal years, the estimate for the 'rate of change' will alter in value quite substantially. For example, if the final value in Table 5.9 had been 60 and not 63, the rate of change would have decreased in value to 5.0 per cent. The estimate calculated in this way is thus quite unstable and susceptible to major fluctuations.

The value calculated also assumes a constant increase each year – that constant being calculated as a proportion of the initial value. Hence computation of values for intermediate years, when this is necessary, is effected according to the formula:

$$V_i = I + ROC.\, t \qquad \textit{(Equation 5.15)}$$

where I and ROC are defined above
t = years from initial year.

This formula implies a different concept to that normally attached to the interpretation of 'rates of change'. Rarely, for example, are baseline values used throughout a period for estimation purposes. More frequently, percentage changes apply to each successive year. Partly for this reason, therefore, estimates of rates using Equation 5.14 are not usually made.

A geometric mean value estimates a multiplicative rate which is more frequently used – at least in economics – than is the additive model described by the last equation. Here the 'rate of change' is calculated as follows:

$$ROC = \sqrt[N-1]{\frac{F}{I}} - 1.0 \qquad \textit{(Equation 5.16)}$$

where the symbols are as for Equation 5.14.

The value calculated by this method for the data in Table 5.9 is 5.95 per cent. Again the same weakness as was noted for the overall arithmetic average applies. Variation in the value for either the initial or the final year significantly affects the estimated 'rate of change'. The prediction model here is multiplicative not additive. It can be summarized as:

$$V_i = V_1 \left(1 + \frac{ROC}{100}\right)^t \qquad \textit{(Equation 5.17)}$$

This equation is much more commonly used than is Equation 5.15 based on the arithmetic change.

Another way of estimating 'rate of change' is to account for intermediate values and to calculate the arithmetic mean of these annual changes. The estimating formula for this method is:

$$\mathrm{ROC} = \frac{\sum_{i=1}^{N-1}\left[\dfrac{V_{i+1}}{V_i} - 1\right]}{N-1} \qquad \textit{(Equation 5.18)}$$

The value estimated by this equation for the data in Table 5.9 is 6.27 per cent which is about midway between the other two estimates. This method does at least take into account values for each intervening year in a period and it thus forms an estimate of the value of the 'most likely change' (in statistical terms). It tends to under-estimate change when there have been more increases in value than decreases while it over-estimates change when there have been more decreases than increases. Estimation of intervening values can again be made by using Equation 5.15.

There is no point in trying to estimate a geometric mean value from intermediate values in this way. The value which would be estimated is exactly the same as is calculated by Equation 5.16. This is because the geometric mean method implicitly cancels out all intermediate changes so that only the first and last values really enter into the calculation. This method of estimation is thus 'path independent'. Sometimes this provides a stable estimate for the 'rate of change' but it requires very good estimates for the values of a variable in the initial and final years.

A final method for estimating change rates to be discussed here is simple linear regression. This method attempts to fit a regression line through the observed values for a variable over the period being investigated. This requires substitution into the appropriate formula[1] of pairs of numbers such as those in the first two columns of Table 5.9. After calculating the mathematical constant values, an equation of the following type is formed:

$$VV = a + b.t \qquad \textit{(Equation 5.19)}$$

where VV = value for variable

---

[1] Most basic statistical books provide discussion of simple linear regression. The computing formulae for the two coefficients in Equation 5.19 expressed in terms of the present situation and variables are:

$$b = \frac{N \sum_{i=1}^{N} V_i . t_i - \left(\sum_{i=1}^{N} V_i\right) . \left(\sum_{i=1}^{N} t_i\right)}{N\left(\sum_{i=1}^{N} t_i^2\right) - \sum_{i=1}^{N} (t_i)^2}$$

$$a = \bar{V} - b.\bar{t}$$

where V_i = value for variable in year i

N = number of years in period over which the regression line is being estimated

$\bar{V}$ = average value for variable across period

$\bar{t}$ = average of all years in period.

a = constant (estimated by the computations)
b = amount of annual change (estimated by the computations)
t = year number.

The change coefficient (b) for the independent variable of year (t) gives the estimate of the amount of change which occurs in the dependent variable each year. For the situation described in Table 5.9, the equation is calculated to be:

$$VV = 46.9 + 2.5\ t.$$

Hence, for each year a change of 2.5 units occurs in the value of the dependent variable. Expressed as a percentage of the first year value, this gives a growth rate of 5.00 per cent. As Diagram 5.4 shows, the regression line does not actually pass through any of the observed values – not even the first or the last. Indeed it is well below the other lines drawn. What the regression line attempts to do is to estimate the line which best predicts the location of all points. In any given situation, the extent to which the line does this is reflected in the correlation coefficient between the dependent variable and the independent variable (ie between VV and t). Sometimes the fit obtained can be excellent; at other times, however, especially when major fluctuations have occurred, the fit is not very exact. Despite this, the regression estimate is still a most preferable one because it attempts to account for all observed values.

All five methods discussed provide different estimates for the rate of change in this example. The estimates range from 5.0 per cent for the regression method to 6.5 per cent for the overall arithmetic average. The immediate implication of this range is that the least error-inducing method must be chosen wherever possible and adhered to. If it is not, indicator values will show a change merely because of the different methods of calculating values for the components, and this is not a desirable situation. Generally, for measuring change in education systems, the method of regression is to be preferred because it relies on information from each year over the period being examined and it treats all values as being equally important. Its only real weakness is that it is based on the additive model and not on the multiplicative model. The next preferable methods are the geometric mean method and the annual arithmetic mean method.

Separation of male and female data

A second problem concerning the technical aspects of computing values for the indicators recommended relates to whether total human resource variables should be used or whether the components of these variables – males and females – should be reported separately. For example, input indicator 1 requires as one of its variables the net first-level enrolment ratio. This value can be reported for all first-level enrolments together or it can be reported as two variables – the ratio for males and the ratio for females. It is recommended that, wherever possible and appropriate, values for variables should separate males from females. When this is possible, much greater insight into the operation of an education system is provided and more precise monitoring of system development is allowed. Many education systems data do not, unfortunately, separate the sexes, and for these systems overall data only can be used.

This observation leads on to the problem of coping with missing data. This problem is a crucial one especially when cross-national comparisons are to be made. Many data, even of the elementary type referred to in the tables listing the basic indicators, are not available in many countries – especially in some of the least developed countries. In order to make good estimations, therefore, indicator values should, whenever possible, be calculated as averages over as many component variables for which data are available. When indicator definitions are based on the Gini coefficient or the C index, averages are not appropriate. Hence, nothing can be done to estimate an overall value for these indicators if the component variables do not have data available for them. For some of the indicators – for example numbers 1 to 6 in the process set (Table 5.4) – averaging a fewer number of variables than is desirable would at least provide an approximate estimate of a value on a given indicator. Such a procedure overcomes the problem of being forced to use total enrolment data in one situation but the preferable separated male and female data in another situation or for another country.

Values for different indicators need not be comparable
A third problem may be seen in the differences in the range of values estimated by the various recommended indicators. Some values, for example, range between 0.0 and 1.0 while others could have a range of 30 to 50 points at least. This is not really a difficulty because each indicator must be interpreted in light of the criteria which are appropriate for it. Hence the standardization adopted by some writers (eg Unesco [1974]) to ensure a constant range for all indicators in a set, is not considered necessary here. Certainly for some research purposes, indicator values can be standardized, but in most planning and monitoring situations, such a procedure is not necessary. Some of the discussion in Chapter 7 elaborates ways in which indicators might be presented graphically and, on some occasions, visual equality might be desirable. Ways to achieve this are discussed later.

5.6 Résumé

A number of indicators for use in the study of national education systems can now be recommended to planners and researchers. Some of these indicators have evolved from empirical studies while others have

202

simply appeared to be concerns of importance. Certainly additional concerns can be thought of which are probably of equal importance. The present proposals are, however, intended to form preliminary lists which might be regarded even as the common cores of basic sets of education system indicators for each part of the defined systems space. Research and development should now be undertaken to analyse these sets of indicators and to provide hard evidence to justify the adding to or deleting from the components in each set.

In reviewing studies for inclusion here, a strict adherence to the definition and concept of indicators proposed in Chapter 1 has been observed. If this had not been the case and studies suggesting individual variables had been cited as well, a plethora of studies would have been included. The result of observing the criteria is the revelation that there has been and still is a dearth of coherent work conducted in the field of education system indicator formation. That this is so is a pity and it highlights an area to which planners and researchers should devote considerable energy in the immediate future.

It should again be pointed out that no normative guidelines are recommended here. Values incorporated in such guidelines should derive from theoretical statements and from empirical observations. Neither of these is available in sufficient quality to allow critical values to be identified. In addition, as is maintained in Chapter 1, it is only when an indicator value is related to the situation and the system from which it is derived that an interpretation of a normative kind can be made.

Chapter 6

Indicators Developed for Regional Education Systems

6.1 Characteristics of regional indicators

Education sub-systems exist below that of the national system in a country. The next and most immediately obvious level at which education systems are provided is that of the state or province in a country. With increasing attempts being made to decentralize educational administration in both developed and developing countries alike, indicators based on data describing state education systems become increasingly more important. Below this level there are other stages as well – such as local regions or areas, inspectorates and so on down to the level of the individual school itself. At this last level, indicators developed within the context of educational planning can very often be classified just as properly into other disciplines as well – the classroom interaction area, educational psychology and educational sociology being the main three contenders. It is for this reason that indicators developed by studies focusing on the operation of a school must be excluded from consideration here. If this is not done, very many indicators would need to be considered although few of them would have direct relevance to educational planning. Indicators such as the sub-scales of Halpin and Croft's (1965) Organizational Climate Description Questionnaire (OCDQ) are thus excluded as too are the numerous community involvement scales (eg Finlayson *et al*, 1971). Although values obtained on these indicators could be used to guide policy decisions, they are not really planning indicators as such.

In recent years, increasing attention has been given to planning and research at the local level. Reasons for this are numerous and very often specific to a given country or entirely political or a combination of both. In the US, for example, the increase in attention to educational planning at the local level can be attributed to the Title III assessments now required by law of local school districts and of the states themselves. Partly also the increase in attention can be attributed to the worldwide accountability movement. For too long education has been regarded as a special sector into which funds can be constantly channeled. With increased cost and inflationary pressures being exerted on governments around the world, the education section can no longer continue to accept such large sums of money without demonstrating its growth and improvement in all aspects. So that some specificity can be introduced into any debate, state and regional education systems are being increasingly used as the unit of analysis.

The adoption of such a unit as this also has an immediate statistical advantage. The geographical area covered by the data is much more limited than when a total national system is examined. It is thus easier to collect and process data on regional systems. Assuming normal precautions are taken, those data should be much more valid and reliable than data estimated for an entire nation.

Despite these obvious advantages in using education systems within or for a province or a region as the basis for developing indicators, very little development work is available in published form. This is probably partly attributable (paradoxically) to the factor noted above – that detailed and very reliable data are available at this level. Because of this, many tabulations are published for very specific variables. For example, from the regional and the individual school level in many parts of the world, statistics of the percentages of final year secondary students receiving certain kinds of scholarship are often published. As the data are available in such specific form, an attitude has prevailed that aggregating them would destroy their precision and therefore their usefulness. Certainly for some planning functions such an attitude is very correct. For many planning purposes, however, aggregating variables to form proper regional education system indicators allows far greater insight into the operation and development of these systems.

A second reason for the unavailability of reports on indicator development at this level is that, although the area has received attention in some regional education offices, many of the reports prepared are confidential and therefore cannot be referred to here. The imposition of these restrictions impedes the flow of information about the usefulness of certain experimental approaches to indicator development. In many cases, this confidentiality is certainly an understandable restriction. Often, for example, funds are allocated to schools on the basis of these indicators. Releasing the technical details would increase the unreliability and/or invalidity of data reported by those schools wanting more funds. In other circumstances, however, the imposition of a confidentiality marker is simply to prevent discussion on methods of fund allocation. Sometimes unfortunately only a single, very imperfect variable – such as the teacher-pupil ratio – is used as a basis for such decisions. This limited availability – for whatever reason – of good discussion papers describing the development of indicators for regional systems forces emphasis in this chapter on to a limited range of developmental efforts.

Almost without exception, indicators developed at this level, in both published and unpublished studies, focus on either the process or the output sub-divisions of the education system space. This is a generally expected concentration. It is traditional, for example, to focus most attention on achievement test scores and/or graduation rates. These have long been considered not only important outputs of an education system but also the most important variables to reveal how a system is functioning. Consideration of processes has lately emerged because of the need to distribute specific grants to schools and these grants have been aimed toward special school operations and characteristics. Because of this concentration on two parts of the education system space, only process and output indicators can be discussed in this chapter.

Before discussing the indicators in these two sub-divisions, however, the set of input indicators recommended for national systems must also be recommended for adoption and application to regional education systems. Naturally appropriate redefinition of some variables must occur – especially to those variables forming the financial involvement index. The use of these indicators at a regional level is a different concept in many ways to that which has generally prevailed to date. Nevertheless, it underlines the fact that it is important to consider the involvement of the human and financial resources of regional communities in regional planning and policy formation. This importance immediately suggests that the measurement and evaluation of particular characteristics is a *sine qua non* at this level and that input indicators can serve an equally valuable function at the regional level as they serve at the national level.

For the process and output sub-division, the same indicators as are recommended in the previous chapter are recommended also for regional education systems. As with the input indicators, redefinition of some component variables must occur. This recommendation is taken after again noting that there are many features of education systems at the lower levels of aggregation in both of these sub-divisions which do not differ in their basic nature from those characteristics present in national systems. This adoption thus makes a sizeable core of indicators available for immediate consideration in practical planning situations.

If these indicators were the only ones considered for adoption, however, an implicit assumption would have been made that there are no differences between national and regional education systems. As is argued in Chapter 1, patterns of development should be investigated from the individual school level to the national level so as to determine what differences do really exist. If no differences are identified amongst the different levels, then accepting identical indicator sets for all levels is most proper. Such similarities need to be demonstrated in a number of countries, however. There is certainly considerable evidence to suspect that special indicators are needed to describe adequately the unique characteristics of regional systems. Hence, in this chapter indicators additional to those nominated in the context of national education systems are described and some of these are recommended as additions to the basic sets.

6.2 Indicators of regional education system processes

Indicators measuring the processes of regional education systems have tended to follow one of three methods of approach. They have also tended to pursue similar concerns. Indeed three examples cover most of the main features of the approaches used to develop indicators at this level as well as the results which have emerged from their attempts. To discuss all relevant studies would add nothing to the substance of the discussion and would particularize to specific localities unnecessarily. Hence the three studies are now reported in some detail.

The quality of Indonesian secondary schools

Several process indicators defined empirically on individual schools are identified in an investigation conducted in the ministry of education (BP3K) in Indonesia in 1976. One part of this study aimed to define indicators with which to rate the quality of secondary schools. The data to be used for the indicators had to be derived from the annual school returns.

The analysis began by defining a number of variables from the data available. For example, the number of hours taught per week by each teacher in a school was available. A variable measuring the average number of teaching hours per teacher per week can thus be defined for each school. This enables comparisons among schools to be made independently of the size of the schools or the number of teachers located at each school. Analogous ratios could also be formed from a number of variables including some measuring each school's income and expenditure. Some of these variables can be noted in Table 6.1.

As data were available for over 2000 schools, replication analyses were attempted to determine if the indicators identified are stable over different samples. If the indicators are stable, then greater confidence can be placed in their definition. If, however, different indicators emerge from the analyses performed on different samples of schools, it must be concluded that the characteristics identified in each analysis are specific to a particular sample. Attempting to generalize these characteristics to other samples is thus invalid. To undertake these replication analyses, the total sample was randomly split into three equal sub-groups. Parallel independent analyses were then performed on each sample.

Correlation coefficients were calculated for each of the three subsamples across as many pairs of variables for which data were available. Although this is not a wholly desirable practice, eliminating either variables or schools not having complete data would have led to extremely biassed residual sets of variables and schools remaining to be analysed. The effect of forming correlation coefficients across pairs of variables on which not all pairs for each entity being analysed have data available is to destroy the Gramian properties of the correlation matrix. This can sometimes lead to imaginary factors being extracted in a factor analysis. However, a sufficiently large number of pairs of data were available to determine each correlation coefficient and the effect of allowing for missing data in this way is therefore considered to be minimal.

A principal components factor analysis[1] was applied to each of the three correlation coefficient matrices. Eight factors, accounting for between 48.1 per cent and 48.7 per cent of the variance were extracted in each analysis. Only the first four factors were retained for indicator development. The remaining factors reflect the ways in which the correlations had been estimated and they thus identify characteristics particular to certain types of schools.

Table 6.1 summarizes the three rotated factor matrices. Only loadings above 0.45 are reported. The first factor describes the manpower and financial resources of a school. High factor scores on this factor identify schools offering general education in particular and which have a large proportion of the teaching staff who are permanent. They also have a relatively large number of ancillary staff and a high dependence on government funds which are to be used especially for salary payments and the purchase of writing material and equipment. Such a school appears to be a well established school in the Indonesian environment and one which should not have excessive difficulties in its operation.

The second factor describes the quality and the quantity of resources available within a school. High quality schools are identified as having a well qualified headmaster and teachers and good quality facilities. The third factor describes schools with a good environment. Pupils are in small classes which are not crowded and they receive considerable teacher attention. The fourth factor identifies other schools in which teachers are not heavily involved in teaching and they tend to teach to only small classes.

The definition of these four factors is relatively constant across the three sub-samples. As Table 6.1 shows, the factor loadings for each variable are mostly very similar. Where there are exceptions to this constancy – for example variables 5 and 9 in the first factor – a variable which cannot be classified as being entirely reliable for indicator formation is identified. For the formation of the required indicators of quality, these variables were not used.

The values for each school on those variables chosen were standardized and then combined additively with equal weights. Averages were

---

[1] The program used is BMD085. Communality estimates were squared multiple correlation coefficients. Varimax rotation was performed on the eight factors with eigen values greater than 1.0.

	No.	Variable Name	Factor loadings from sample of schools		
			A	**B**	**C**
Factor 1: Manpower and financial resources of the school	1	% of permanent teachers	0.88	0.85	0.88
	2	total salary and honorarium per teacher	0.62	0.58	0.68
	3	Average teacher hours per week per student	0.83	0.79	0.82
	4	Offering of general education	0.89	0.88	0.91
	5	Number of ancillary staff	0.46	0.41	0.50
	6	% expenditure from private sources	−0.91	−0.93	−0.93
	7	% expenditure from government	0.94	0.96	0.95
	8	% honoraria from government	0.68	0.69	0.70
	9	% expenditure on teachers' salaries	0.56	0.52	0.63
	10	% expenditure on honoraria	−0.56	−0.52	−0.64
	11	% expenditure on equipment from government	0.49	0.53	0.50
	12	% teacher salaries from government	0.87	0.88	0.90
	13	% salaries of ancillary staff from government	0.84	0.84	0.84
	14	% expenditure on writing from government	0.80	0.81	0.77
	15	Expenditure from private sources	−0.47	−0.53	−0.51
Factor 2: Quality and quantity of resources in a school	1	% teachers with PGSLP (Teacher training diploma for lower secondary school leavers) qualifications	0.54	0.51	0.54
	2	% teachers with a degree	0.75	0.73	0.79
	3	Education level of headmaster	0.55	0.54	0.53
	4	Facilities	0.54	0.56	0.56
	5	Number of streams	0.61	0.62	0.68
	6	Total expenditure on writing materials	0.50	0.51	0.42
	7	Total expenditure	0.62	0.67	0.54
Factor 3: Pupil environment	1	Lesson hours per student per week	0.79	0.81	0.79
	2	Classroom area per student	0.62	0.57	0.63
	3	Students per teacher	−0.71	−0.68	−0.71
	4	Class size	−0.85	−0.87	−0.87
Factor 4: Teacher involvement	1	Average teaching hours per teacher per week	0.85	0.89	0.85
	2	As above for permanent teachers only	0.91	0.87	0.88
	3	Classes per teacher	0.84	0.86	0.85
	4	Students per teacher	0.58	0.63	0.56
		% of variance accounted for by each factor 1	16.2	15.7	16.8
		2	8.5	8.8	7.6
		3	6.3	5.8	6.6
		4	4.9	5.0	4.8

Table 6.1: *Summary of factor matrix for Indonesian school quality analysis*

then calculated across those components which were available to be used.

The indicators defined in this study appear to have face validity in that intuitively the first three factors especially seem to be useful for identifying schools with high quality environments. Of course, none of the indicators examines considerations of quality in the teaching-learning process because of the nature of the exercise undertaken. To that extent, they do not offer a complete description of the education process within a school. It is very likely that the indicators identified here could be generalized to other regions both inside and outside Indonesia.

The identification of under-privileged London schools

Another study developing a process indicator is that by Little and Mabey (1972). This describes an attempt by the Inner London Education Authority to develop an index identifying 'those schools which need special help' (1972, 67). The background to the problem is provided by the Plowden Report which introduced the concept of 'educational priority areas'. The report does not however define this term precisely. The index discussed by Little and Mabey attempts to quantify the concept. Their project is therefore 'an attempt to create an instrument that turns a policy objective into administrative practice' (1972, 93). As is the case in so many similar situations, administrative constraints were imposed on the free development of the indicator as it had been originally conceived. Not only did the index have to fulfil the Plowden concept but it was also to be used as a guide for the funding of particular schools. Hence not all variables used are necessarily the ones which might have been defined.

The basis for the definition of the index was provided by Little and Mabey analysing, in an *a priori* way, the concept implied by the Plowden Report and the administrative intent of the ILEA. They then assessed what they had identified as being required against the extent of the availability of data which could be accessed from the usual annual school returns made to the Department of Education and Science as well as from census information. Ten variables were thus identified to form the components of the indicator. They are listed, together with their operational definitions, in Table 6.2. The final composite index is calculated for each school by standardizing each variable according to Equation 3.3 and summing the equally weighted values. All ILEA schools were used to compute the basic statistics used for the standardization.

Little and Mabey attempt to investigate the validity of the index by actually inspecting some of the schools. They choose schools from different parts of the indicator distribution and consider, on the basis of their inspections, that the indicator does have sufficient validity in its identification of schools needing special help to justify its use.

Variable	Operational definition
Social class	% employed males in district in unskilled or semi-skilled jobs
Family size	% children living in households containing 6 or more people
Overcrowding	% households in district living in a density of more than 1½ persons per room
Housing stress	% of households in district without an inside lavatory
Free meals	% of pupils in school receiving free meals
Absenteeism	Average absences from school during a nominated week
Immigrants	% of pupils in school who are immigrants
Retarded or handicapped	% of pupils scoring at or below the 25th percentile at the 11+ transfer stage on a NFER standardized test
Teacher turnover	% of full-time teachers in the school who had taught at that school for less than 3 years
Pupil turnover	% of pupils attending the school for less than one year.

Source: Little and Mabey (1974)

Table 6.2: *Summary of variables used for each school and its school district*

They claim to have considered factor analysis as an alternative to their theoretically based approach to the development of the indicator but discarded this because the analysis would not have had sufficient explanatory power in their situation.

The indicator is a complex one to interpret. It combines variables related to so many different aspects that any value for the index could result from many different patterns in a community. Indeed the indicator suffers to some extent from over-aggregation. In addition, it goes beyond a dependence on variables related only to the education system and incorporates variables describing social conditions as well. Primarily for this latter feature, therefore, this index cannot be recommended for adoption for the purpose of planning, monitoring and evaluating education system development.

A curriculum index

The third example of a process indicator[1] applicable to regional education systems is provided by Barrows and Shaffer (1975). They develop an index which primarily measures the kind of curriculum being implemented in the US state of Wisconsin. The indicator is identified through factor analysis and it is composed of the variables reported in Table 6.3. The four curriculum variables dominate the index quite substantially. Whether this is really the case in the actual situation or whether the apparent dominance is caused through overloading the variable space to be analysed with curriculum variables can only be surmised. Further definition of the index is given by the 'subjects per teacher' variable which has a very high negative loading.

---

[1] Although the indicator is classified by Barrows and Shaffer as an input indicator, it is reclassified here as a process indicator because of its nature *vis-à-vis* the definition of a process indicator given in Section 1.5.

The Barrows and Shaffer index is difficult to interpret from a planning point of view except to name it a 'variety of curriculum' index. The use the researchers make of it is interesting, however. They provide a map of the state of Wisconsin divided into counties. On this are indicated the ten counties with the highest indicator values. These counties are predominately in the south-east corner of the state or they are urban counties. There is a discernable bias therefore in the location of the counties offering such a variety of curriculum as is indicated by this indicator. Hence the addition of such a map as Barrows and Shaffer provide can help to ascertain where particular action might be necessary to equalize opportunities for students.

Variables used to construct index	Rotated factor matrix coefficient
Net operating cost per average daily admissions (1970)	.038
Specific state aids per average daily admissions (1970)	.640
Subjects per teacher (1970)	−.774
College prep curriculum per school district (1970)	.824
Business curriculum per school district (1970)	.734
Vocational curriculum per school district (1970)	.748
Cultural curriculum per school district (1970)	.869
Percent of total variance explained = 50.6	

Source: Barrows and Schaffer (1975, 350)

Table 6.3: *Variables defining the Barrows and Shaffer education index I – Wisconsin, USA*

The index as defined is specific to Wisconsin. It probably has limited generalizability outside of that US state because of the nature of the curricula defining it. These curricula are not always present in a school system; nor are they necessarily measured in that same way when they are. For these reasons, the Barrows and Shaffer indicator cannot be adopted as it has been defined for any basic set of education system indicators. The index does however represent a most important concept which should be measured by a basic indicator. Planners must therefore be encouraged to suggest possible definitions for the indicator which might be generalizable across regions and even across nations. Any such indicator, if it were to be defined, should complement the indicator measuring the orientation of second-level students. This latter index indirectly measures the 'variety of curriculum' offered at the second-level. A direct measure is required, however.

One way in which a direct indicator might be developed is to estimate the proportions of total school time devoted to particular types of content or to particular types of activity. A C index of concentration (Equation 3.10) calculated from these proportions would comment on the variety in the curriculum or of the curricula offered at a particular level. The validity of this suggestion needs to be tested empirically. Until it is, an indicator measuring this important area of the process sub-division of the education system space cannot be included in a core of planning and monitoring indicators.

6.3 Indicators of regional education systems outputs

A limited variety of studies have so far attempted to measure the outputs from regional education systems. Some of the work that has been conducted is reviewed here and it should be considered as supplementing the indicators recommended in Section 5.3 for the measurement of outputs from national education systems. Further supporting this recommendation is the observation that although three broad areas were identified in the previous discussion of outputs, only one area – that of achievement levels reached by students in or leaving the education system – has received attention at the regional level. Two distinct approaches have indeed emerged within this area – one aimed towards the actual undertaking of surveys to obtain achievement data, the other oriented more towards analysing the data which are collected.

Surveys of achievement level

Numerous examples of surveys conducted to assess levels of achievement can be cited. Many countries have now realized the importance of measuring the achievement levels of their students in a coherent, objective and standardized way. Before such recognition, standardized tests of various types were used to measure certain skill levels. The results of these were considered to be useful mainly within schools, and not for assessing achievement across schools in a region or province. Tests such as these in the Iowa Basic Skills Program in the US and the NSW Basic Skills Testing Program in Australia are typical examples of tests of this type. Also classifiable here are the Scholastic Aptitude Tests and the Graduate Record Examination in the US.

One major problem with testing programs such as these, at least from the point of view of developing indicators, is that the test items are generally oriented to specific topics in set curricula. The skills tested are those which a specific education office considers to be important for schooling. They are not necessarily skills which can be ranked equally important for functioning in the world outside the school – although many of them are important there also. Given a change in education area, different skills could be considered to be more important. Additionally, a non-representative sample of students attempts these tests each year and this constricts the generalizability of the results even further. Hence there is little basis for comparisons between one region and another with the results from these types of special purpose tests.

One project which has attempted to measure educational achievement across various subject areas and to include representative samples from a variety of regions is the National Assessment of Educational Progress (NAEP) in the US. This project was initiated in 1964. The first national test administrations were conducted in 1969. The project envisaged testing in ten subject areas over a period of about five years. The subjects selected were: art, career and occupational development, citizenship, literature, mathematics, music, reading, science, social studies, writing.

After the initial cycle of testing was almost complete, the re-administration of revised tests in each subject area was planned. These revised versions were to contain a high proportion of previously used items. The second, and in some cases the third, administration of tests in most subject areas have by now either taken place or are to be completed shortly. For each test, four groups of individuals are identified. They are the 9 year olds, 13 year olds, 17 year olds and young adults (defined as people in the 26–35 year range).

In the details discussed to this point, there are only minor differences between the NAEP and the IEA studies referred to in Section 5.3. What distinguishes the two projects sufficiently to allow their different classification is the sampling scheme used for each of the four age groups in the NAEP. In the IEA project, a national sample is taken from the countries participating in a particular subject testing session. Very few results about areas within those countries can therefore be validly drawn. In the NAEP study, samples are drawn within each age group in particular areas of the US and from people with specifically defined characteristics. The sampling grid adopted stratifies the population into:

☐ four main geographical regions
☐ four different types of community
☐ two socio-economic levels
☐ males and females.

People to whom the tests are to be administered are drawn within each of the cells formed by the four dimensional grid. Certainly the coverage effected by this grid does not include all parts of the US. It does however facilitate a discussion of many regional similarities and differences on the basis of the NAEP results.

An additional difference between the NAEP and the IEA study is that the tests in the former are constructed and administered by employing matrix sampling procedures.[1] These basically require a two dimensional matrix to be formed which has 'people' as one axis and 'test items' as the other. From the matrix so formed, a sample of people is identified as well as a sample of test items. These items are then administered to those people. A second sampling of people and test items is then made, then a third sampling and so on. Hence not all examinees complete the same set of questions. Comparisons among groups of individuals with particular characteristics must therefore be made with care.

The national assessment results do allow some analysis of regional variations in educational progress. Certainly more specific comparisons do sometimes need to be made. When this is so, subject results in nationally administered examinations can be used. Unfortunately, in many countries, these examinations do not provide a good basis for these comparisons. Their validity and reliability are often questionable, they generally have a very specific content domain, there is a lack of questions common from one administration to another and the standardization procedures applied to transform the raw marks nullify anything but a superficial, norm-referenced comparison among regions for any single year.

Some of the other national surveys of educational achievement mentioned in Section 5.3 can, like the NAEP, be used to examine regional variations within nations. Indeed many of the reports published do compare — at least descriptively — one region to another on sub-test scores and/or on total scores. Such scores can be regarded as being elementary indicators of a kind, although they are conventionally regarded more as individual variables. In either classification, sub-test scores must form a basic component in any set of measures designed to compare among education systems in regions within a country. Total scores are not as preferable as sub-test scores because their excessive aggregation over so many content areas and cognitive levels makes the interpretation of specific achievement scores extremely difficult. To ensure widespread applicability and understandability of the results, more achievement tests should be constructed which break away from specific curriculum-based content and skills. They should, like the NAEP, attempt to assess cognitive abilities and skills which are considered by leaders in a society and by educationalists as being important for functioning in that specific society. Such a move will in turn lead to an increase in the efficiency of monitoring achievement by those test administrations.

The analysis of national survey data

Researchers who use the data collected in surveys and national test administrations are also becoming aware that indicators of some sort should be constructed and used to monitor educational progress especially at the district or regional level as well as at the individual school level. Some of the impetus for this movement has stemmed from the need to evaluate both the effectiveness of special programs and the benefits conferred on individual schools by special grants. Typical of the approaches being pursued here are those by Dyer *et al* (1969) and Marco (1974). They define and test various indicators of school effectiveness which appear to them to be intuitively useful. Marco, for example, uses longitudinal data collected from 70 midwest elementary schools in the US which participated in a testing program in both 1970 and 1971. From these schools, 3769 students attempted tests in both years. The two test results for each of these students were thus used to devise school effectiveness indices.

---

[1] For a discussion of the concepts and mathematics of matrix sampling see in particular Husek and Sirontnik (1968) and Shoemaker (1975).

Four of Marco's indicators are based on a regression line of the type:

$$Y_{1971} = a + b\,X_{1970} \qquad \textit{(Equation 6.1)}$$

where Y_{1971} = the 1971 test result

X_{1970} = the 1970 test result

and a and b are parameters estimated by the regression calculations.

This line can be estimated in different ways according to the group being analysed and the level of aggregation of the test results. Some of the indicators defined by Marco are based on the estimated values for the 1971 test results (as opposed to the actual results); others are based on the difference between the actual and the estimated 1971 test results (ie on the residual or the error in the prediction). The actual definitions of the four indicators based on Equation 6.1 are:

Indicator 1: the predicted 1971 scores for low, middle and high scoring students as estimated from the 1970 results. The regression equation is calculated for each school independently on the assumption that 'a school may be more effective for one type of student than another' (1974, 227).

Indicator 2: The same method as for indicator 1 but with test results for each year being corrected for their unreliability.

Indicator 3: The average residual scores for a school. The regression equation is to be calculated across all students from all schools.

Indicator 4: The same definition as for indicator 3 but with the regression equation calculated across all schools by using only average 1970 and 1971 test results for each.

The final indicator is not based on regression and is very simply defined:

Indicator 5: The difference between the mean 1970 and 1971 test results calculated for each school.

Marco analyses the inter-relationships amongst the indicators as well as their stability. He concludes that although the indicators are highly inter-correlated,

> they are different enough from one another and in their relations with other variables to prevent them from being used interchangeably. The school effectiveness indices for initially low- and high-scoring students appear to yield unique information and raise doubts about using a simple index to measure school effectiveness. (1974, 233)

The results of these explorations into data analysis point out one direction which planners and researchers should be following if test results are to be used for more than descriptive comment on school effectiveness. At this stage, Marco's results are admittedly tentative and not conclusive. For this reason, none of his indicators can be recommended for adoption on a wider scale.

What can be recommended on the basis of the results of Marco's analyses and from the NAEP re-administrations is that planners ensure they undertake or be associated with large-scale assessment projects which envisage re-administrations of the testing instruments in future years. If data are only available for one year, thorough evaluation is not possible and the task of developing indicators is made superfluous. Longitudinal data must be available if useful and proper evaluations of education system development are to be attempted. Additionally, it is only through being able to access longitudinal data that reliable and valid indicators of this kind of education system output can be developed.

Levels of educational attainment

A final indicator developed for measuring the output from regional education systems measures educational attainment levels. It is proposed by Barrows and Shaffer (1975) and emerges from their factor analytic study of the US state of Wisconsin referred to in the previous section. The variables defining the factor are reported in Table 6.4 together with their factor loadings. They measure the amount of education already received by members of the population in each region of Wisconsin as well as the percentage of the older school age population still attending school. Regions with high scores on the index therefore have high educational attainment levels and are encouraging the children in the region to maintain those levels. The index combines variables which are often individually cited as being amongst the most important for measuring the outputs from education systems.[1]

Barrows and Shaffer provide a graph of the location of Wisconsin's counties and mark on it the ten counties having the highest factor scores on this indicator. Those counties with university campuses are

---

[1] Duncan (1968) is one writer who adopts such a stance by resting her entire case on the variable measuring median years of schooling.

Variables used to construct index	Rotated factor matrix coefficient
Median years of education of males (1970)	.931
Percent of males over 25 years old with a high school diploma (1970)	.945
Median years of education of females (1970)	.891
Percent of females over 25 years old with a high school diploma (1970)	.948
Percent of 16 and 17 year olds still in school (1970)	.381
Percent of total variance explained = 72.0	

Source: Barrows and Shaffer (1975, 352)

Table 6.4: *Variables defining the Barrows and Shaffer education index II – Wisconsin, USA*

clearly identified by these scores.

Except for the variable measuring the percentage of 16 and 17 year olds still in school (which is really an input variable) none of the variables defining the index necessarily comments on the Wisconsin state education system. Population mobility is one factor which would work against such an interpretation. It is unlikely, for example, that the educational levels of people immigrating into the state would be exactly balanced by the levels of those people who were emigrating. The relationship between these levels would be particularly complex in those counties having university campuses – and on which the highest factor scores were calculated. Certainly the index does measure educational attainment levels as presently exist in Wisconsin. The implication that those levels can be attributed to the Wisconsin education system cannot be made, however. As it is with the outputs from specific education systems that indicators should be concerned, the Barrows and Shaffer index cannot be recommended for use in the context of educational planning.

6.4 Résumé

This review of indicator development for regional systems has many constraints on it. It must therefore be much more abbreviated than the review of national education systems. Indeed no indicators additional to those recommended in the previous chapter – with appropriate re-definition of the component variables where necessary – can be recommended for planning, monitoring or research purposes at this level of aggregation. That this is so is a further indictment of the state of development of education system indicators. Indices such as those developed by Marco for measuring school effectiveness should be investigated further. His indicators at least do offer some framework on which additional research could build. Marco's indices, as he freely acknowledges, are not yet so developed as to be able to recommend them for widespread use. It can only be hoped that planners and researchers undertake longitudinal studies on regional systems which might enable specific planning indicators to evolve for this level.

Chapter 7

The Presentation of Indicators

To this point, discussion has been focused on the technical aspects of defining and developing education system indicators. Having now established an appropriate basis for indicators and identified some indicators which might prove to be useful for educational planners and researchers, this and the following three chapters direct attention towards ways in which indicators might be used.

Perhaps the most common use for indicators is the communication of information about what a particular education system is like. This is, of course, also the aim of the multitude of statistics produced each year by every ministry of education although, in this latter context, specific rather than general system characteristics are being communicated. Unfortunately the presentation of statistical information in a proper way is rarely achieved. Indeed there are so many problems here that a whole new area of interest has emerged. This area – called kinostatistics – involves the development of both printed and non-printed materials for the purpose of communicating a statistical message in a clearly understandable way. Some of the principles of kinostatistics are referenced in the following discussion.

A major difficulty to overcome when developing methods of presenting statistical data – especially data in the form of indicators – is that the audience to whom the report is addressed is generally exceptionally diverse. Politicians, administrators in a variety of fields, educationalists, researchers and the public at large have differing degrees of interest in reports on education. Their differences in background as well as in perspective generally ensure that data presentations should be as straightforward as possible and fully explained. Technical jargon and unwarranted assumptions must be kept to a minimum or eliminated entirely.

It is generally agreed that messages about social and economic conditions – and not just about education systems – are not communicated effectively through current statistical presentations. In noting some of the difficulties, for example, Feinberg states:

> Every day the American public receives fragments of a social report on the quality of life in their nation. It takes the form of diverse bits of information, haphazardly strewn about by the government, media, industry and the academic community. From this mass of uncoordinated and sometimes poorly collected and erroneously reported data, citizens are challenged to become knowledgeable enough about the social condition of their society to consider intelligently alternative social policies that might effectively deal with a whole bag of social ills. (1972, 1)

Biederman also notes this problem and posits that the failure to present properly information about general social conditions can be attributed to 'the tendency of the attempts to fall between two antithetical stools: the need for scholarly accuracy and complexity, versus the problems of communicating results to audiences of non-specialists in social measurement' (1971, 14).

Traditionally social statistics are presented through the use of tables. This is by far the simplest approach and perhaps the most common. Fairly complete information can be conveyed in this way and cross-references and explanations can be made where appropriate. On the basis of experimental evidence which compared the ability of people to interpret tables and various pictorial representations of data, Carter concluded that a properly presented table 'is as rapid, and more accurate to use, than any other method of presenting data' (1947, 649). The generally poor use of pictorial representations would probably allow this conclusion to be still supported today over 30 years later.

There are two problems with presenting data only through tables however. One is that many of the procedures which must be observed in presenting tables are often ignored by compilers of official statistics. Incomplete or inaccurate description of the variables or indicators being reported, poor cross-classification by descriptive variables such as region, sex, age, etc and over-aggregation compound more basic yet frequently encountered errors such as totals or percentages not being correct.

A second problem with using only tables is that many readers – both specialists in a given field and non-specialists – mentally 'turn-off' when confronted by a vast array of figures. They do not or cannot make the effort to interpret what message is presented. Sometimes of course, such an attitude is understandable because a compilation may be badly presented or a very complex trend or inter-relationship may be present. In such circumstances, a different mode of presentation is necessary. At other times, tables overload a reader with too much information and it is impossible to reduce this morass to the basic elements of its message. Although non-printed forms offer much in overcoming the limitations of tables as contained in these situations, only static graphical alternatives are outlined here. Most planners, in their reports, will have to rely on only these printed means for some time to come. Even by confining the discussion to this area, however, it is still impossible to attempt to cover even most of the possible methods of presentation. The graphical presentation of data to appeal to the symbolic perceptions of readers is, in some ways, almost classifiable as an art form. There are thus no bounds limiting the development of new methods. When they comment validly and reliably on a set of data, these new methods should be encouraged.

To review some of the main graphical methods, the types of problems to be addressed are identified and these are used as the framework for this chapter. In the first section, methods for reporting data on a single indicator either for one system across a number of years or for a number of education systems for one year are described. This is followed, in the second section, by a description of a method of demonstrating the inter-relationship among a number of indicators. Finally, a method for reporting a number of indicators across a number of systems is presented.

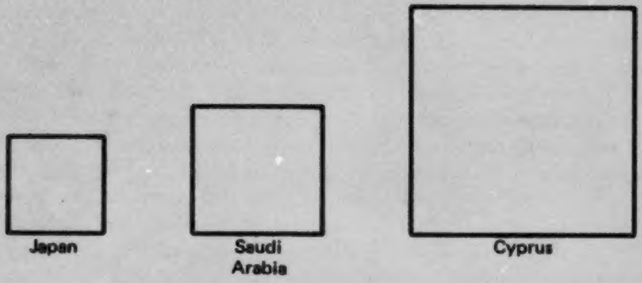
7.1 Methods of presentation involving a single indicator

The presentation of a single indicator in a graphical format is motivated by the need to demonstrate variation in the values of that indicator. The variations in value to be depicted could be within one education system only – either at a single point in time or across a number of years – or they might be variations across a number of systems at a single point in time. Both self-referenced and norm-referenced comparisons can therefore motivate graphical presentations of a single indicator.

Many traditional methods are commonly used to present indicators fulfilling the above purposes. The pie-chart, the square, the cube and the bar-chart are four methods depicting indicator values calculated on a single system or on a number of systems at one point in time. Diagram 7.1 depicts two of these for the data presented in Tables 3.3 and 3.8. All four methods are widely adopted for many kinds of reports and they can be equally applicable to self-referenced and norm-referenced purposes. As they are so frequently used, little explanation or elaboration on any of them is required here. A number of books and articles – Monkhouse and Wilkinson (1963) is one – provide excellent guidance on the proper construction of such diagrams.

One problem does arise in using squares and cubes and also, at times, pie-charts. It is a perspective problem and it derives from the criterion by which an observer is to interpret the visual image seen in a diagram. When length of side is taken as representing an indicator value, a square with sides twice those of another square is really showing an area of four times the original, while a cube drawn to represent a similar difference shows a volume eight times that of the original cube. Attempting to represent indicator values by area or volume is no less difficult because of a confounding on the basis of comparing length of side. For these reasons, the square and the cube are not wholly desirable

(a) Representational squares: percentages of total third-level enrolments in humanities courses, 1971 – data in Table 3.8



(b) Bar charts: female involvement index, 1972 – data in Table 3.3

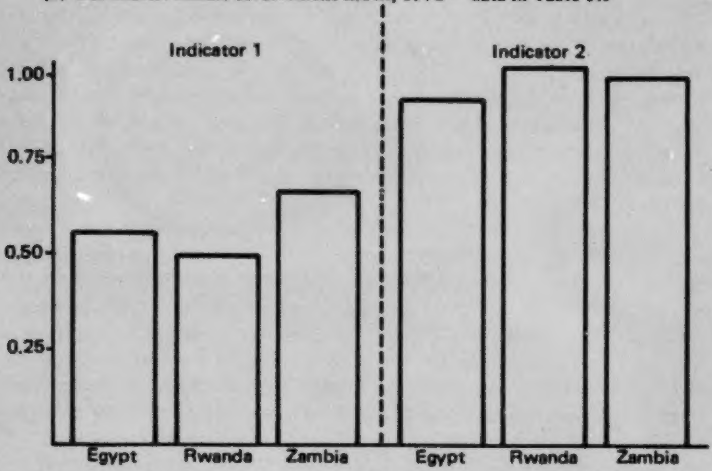


Diagram 7.1: *Examples of different ways to represent a single indicator*

methods of representing indicators.

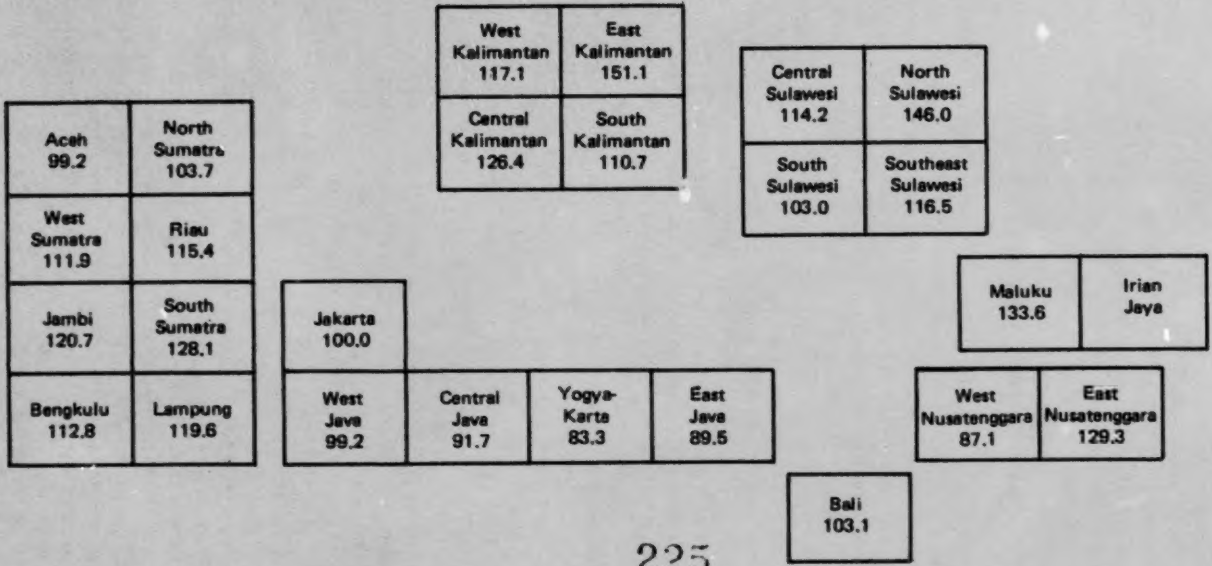
Pie-charts of different sizes can also pose a perspective problem for interpretation. When size of circle is constant, sectors of the pie-charts can be used to compare one percentage value to another. When a number of charts of different areas are presented simultaneously, however, corresponding sectors representing the same percentage have different areas and they thus can convey a different message to an observer.

The bar-chart, as shown in Diagram 7.1(b) has its base values constant. Direct comparisons from one entity to another are therefore possible in terms of either area or length of side. Comparing the interpretability of these various methods experimentally led Croxton and Stein to conclude that 'estimates based on bar-charts [are] more accurate than estimates based on squares, circles or cubes' (1932, 60). When bar-charts are appropriate, it is therefore recommended that they be used rather than one of the other diagrammatic methods mentioned above.

Other ways to present indicators to compare either one region of a country with another or one country with another for a given year have also been used. They depend on the geographical region and they report 'region' as the unit for the indicator value. The simplest way to do this is to write the indicator value on the unit being analysed. Diagram 7.2 presents a typical situation for the regions in Indonesia. There an indicator — the regional price index — is being depicted. Each region for which 1974 values are available is identified by its particular value. Hence those who are familiar with Indonesia can see at a glance the distribution of values and what they imply. Those who are not so familiar can at least gain some insight into the geographical dispersion of the values within the country as well as the locality of the differences in the indicator values. If the same values were to be presented in tabular form, no indication could be conveyed to the reader not familiar with Indonesia as to where, geographically, the differences in values existed. Such a presentation is therefore much more informative than a statistical table both to those who are familiar with the particular regions being analysed as well as to those who are not.

The Barrows and Shaffer (1975) graphs of the US state of Wisconsin, which were referred to in the previous chapter, are further examples of this form of presentation.

An extension of this mode of presentation is to represent world geographic regions by indicator values but to use their 'height above sea level' as the means to represent the values on a particular indicator. Diagram 7.3 depicts a typical graph derived from the World Bank *Atlas* (1979). It depicts the world regional variations in the variable 'GNP per capita'. World regions such as Asia, South America and Oceania are represented by their overall average values on the indicator. These values form the basis for the 'height' of each region. Individual countries could, of course, be represented in this way to demonstrate cross-national variations in the values of a single index. The 'height versus volume' problem noted above for the cube representation is not present here although potentially it may have been. The geographical shape differences are so great that they prevent any perceptual shape or volume comparisons being made. The contours of regional areas and countries are also too well recognizable. Hence the basis for comparison is forced on to the 'height above sea level' distance. For


West Kalimantan 117.1
East Kalimantan 151.1
Central Kalimantan 126.4
South Kalimantan 110.7
Central Sulawesi 114.2
North Sulawesi 146.0
South Sulawesi 103.0
Southeast Sulawesi 116.5
Aceh 99.2
North Sumatra 103.7
West Sumatra 111.9
Riau 115.4
Jambi 120.7
South Sumatra 128.1
Bengkulu 112.8
Lampung 119.6
Jakarta 100.0
West Java 99.2
Central Java 91.7
Yogya-Karta 83.3
East Java 89.5
Maluku 133.6
Irian Jaya
West Nusatenggara 87.1
East Nusatenggara 129.3
Bali 103.1

Source: Arndt and Sundrum (1975, 44)

Diagram 7.2: *Regional price index by province – Indonesia, 1974* both cross-national and cross-regional comparisons, this method of presentation is extremely useful and relatively free of perceptual error.

When the trend in the values of a single indicator is to be traced for one education system over a series of years, an ordinary line graph — properly drawn — is about the most direct and meaningful mode of presentation. This too is a very commonly used reporting form for individual variables. Trends can be easily demonstrated with such a graph as well as the different years identified in which a sharp, atypical rise or fall in the indicator value occurred.

There are a number of variations in using line graphs. When the trend in the indicator value itself is to be depicted, then the graph can be drawn normally with time forming the horizontal axis and the indicator value forming the vertical axis. When the rate at which the indicator value is changing is to be depicted, it is appropriate to transform the indicator values and to plot instead, along the vertical axis, the logarithms of the original values. Using natural logarithms would present a trend such as that estimated by Equation 5.16. Repeating this equation but writing it so that the indicator values are now in logarithmic form, is as follows:

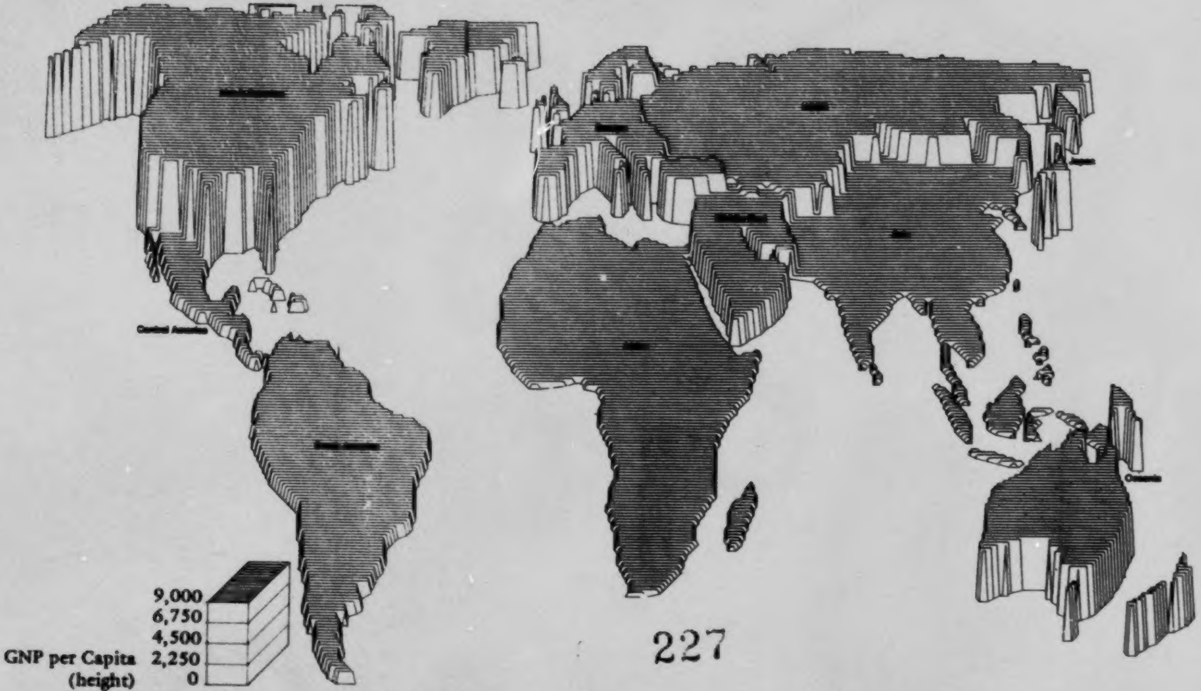
$$\text{ROC} = \frac{1}{N-1} \log_e \left(\frac{F}{I}\right)$$

where ROC = rate of change for logarithmic graph
$\log_e$ = natural logarithm
F = value for the variable (or indicator) in the final year
I = value for the variable (or indicator) in the initial year
N = number of years for which data are being considered.

Diagram 7.4 compares a normal line graph with a semi-log graph to demonstrate the difference being claimed. The calculated values refer to the Academic Production Index discussed in Section 5.3 but for Australian data. The years are 1960 to 1976 and these are graphed along the horizontal axis. In Diagram 7.4(a) the index numbers (standardized to a value of 100 for the base year) are shown while in Diagram 7.4(b) the natural logarithms of the index numbers are graphed. The actual values graphed and the component values for the index are reported in Table 7.1.

The line in Diagram 7.4(a) is shown accelerating at a very rapid rate. This increase reflects the very considerable increase in the total number of degrees awarded at the third-level in Australia over the period being examined here. The line in Diagram 7.4(b) contrasts very markedly with this pattern in that it is very much flatter and does not have the same tendency to increase more acutely in the later years.

The important distinction being reported by these two graphs is that the former comments on the absolute increases in numbers while the latter comments on the relative increases. For example, the line in Diagram 7.4(a) rises most sharply in the years 1965, 1966, 1968, and


Central America
Oceania
9,000
6,750
4,500
GNP per Capita 2,250
(height) 0

The height of each region is proportional to per capita GNP.

Region or country[1]	GNP per capita 1977 (US $)	GNP 1977 (US $000 millions)	Population mid-1977 (millions)
North America	8,710	2,091	240
Japan	6,510	737	113
Oceania	5,490	121	22
Europe, excluding USSR	4,810	2,504	521
USSR	3,330	861	259
Middle East[2]	2,950	130	44
South America	1,360	308	227
Central America[3]	1,120	129	115
Africa	490	209	426
Asia, excluding Japan and Middle East	330	689	2,080
World	1,920	7,779	4,047

NOTE: Due to rounding, the amounts in this table may not equal the amounts or aggregates of the figures appearing in the regional tables.

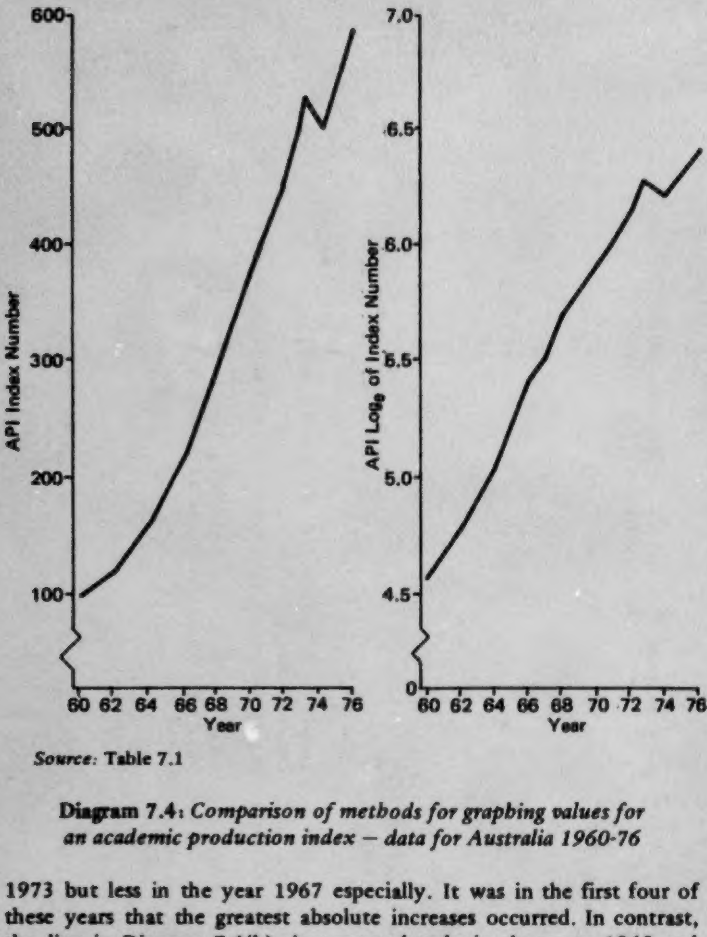
1 Excludes Democratic Kampuchea, Iran, Lebanon, Uganda, and Viet Nam in the aggregation.

2 Consists of Bahrain, Iraq, Israel, Jordan, Kuwait, Oman, Qatar, Saudi Arabia, Syrian Arab Republic, United Arab Emirates, Yemen Arab Republic, and Yemen (People's Democratic Republic of).

3 Includes Mexico

Eckert IV equal area projection

Diagram 7.3: *GNP per capita – by major regions (1977)*
Source: World Bank *Atlas* (1979)



Source: Table 7.1

Diagram 7.4: *Comparison of methods for graphing values for an academic production index – data for Australia 1960-76*

1973 but less in the year 1967 especially. It was in the first four of these years that the greatest absolute increases occurred. In contrast, the line in Diagram 7.4(b) rises most sharply in the years 1965 and 1968 while for the rest of the period it is increasing fairly steadily. It was in both of these years that the biggest relative gains (ie percentage increases) occurred.

Semi-log graphs should be used where it is important to demonstrate relative gain. They are especially useful, therefore, for indicators measuring development emphasis and they can proffer very meaningful comment. Their usefulness in this context would naturally be enhanced if periods of planned development – such as in a four or five year national plan – were also indicated on the graph.

Although variations on all of these methods of presentation are

	Number of degrees awarded			Index number			Composite indicator	
Year	**Bachelor**	**Master**	**PhD**	**Bachelor**	**Master**	**PhD**	**Index**	**$\log_e$**
1960	5021	813	1113	100	100	100	100	4.61
1961	5692	982	1078	113	121	97	110	4.70
1962	6469	1062	1218	129	131	109	123	4.81
1963	7432	1171	1449	148	144	130	141	4.95
1964	8444	1238	1659	168	152	149	156	5.05
1965	9407	1520	2191	187	187	197	190	5.25
1966	10823	1808	2681	216	222	241	226	5.42
1967	12472	2032	2569	248	250	231	243	5.49
1968	13891	2368	3542	277	291	318	295	5.69
1969	16032	2595	3801	319	319	342	327	5.79
1970	16181	2841	4641	322	350	417	363	5.89
1971	17992	3414	5019	358	420	451	410	6.02
1972	20252	3795	5488	403	467	493	454	6.12
1973	23436	4557	6209	467	561	558	529	6.27
1974	25338	4192	5397	505	516	485	502	6.22
1975	26232	4992	5572	522	614	501	546	6.30
1976	27630	5648	5957	550	695	535	593	6.39

Note:

1. Each index number has been calculated using the appropriate 1960 number of awards.
2. The composite index value is the arithmetic average of the three equally weighted components.

Source for columns 2–4: University Statistics (1960-76)

Table 7.1: *Academic production index – values and basic data, Australia, 1960-76*

certainly possible and have been used, the methods described here at least form a basis for other developments and adaptations to particular settings. As the presentation of single indicator values across a number of education systems or for one system across time is one of the most common purposes motivating the presentation of indicator values by graphical means, it is important that further work be done to extend the perceptual basis on which comparisons are presently made.

7.2 Methods of presentation involving the inter-relationships among indicators

A different situation likely to be met in presenting indicator values graphically is where a set of indicators (not just a single indicator) has been defined across a number of education systems. It is then desired to depict the values of these indicators in such a way that it can be demonstrated which values for one indicator are likely to be associated with particular values on another indicator. In order to present results in this way, it is essential to ensure that there is at least a moderate inter-relationship among the indicators being used. Naturally this inter-relationship cannot be too high or else the uniqueness of the characteristics of the systems measured by the indicators must be questioned. If some inter-dependence cannot be established or assumed – as would be the case if a set of empirically defined factors from a factor analysis were to be presented in this way – then any resulting graph would be nonsensical. The graph would merely be a straight line joining the average values of each indicator. The higher the inter-relationships among the indicators, the more meaningful are the lines joining one indicator value to another.

The first major use of this method of presentation was by McGranahan *et al* (1972). Although this study is based on variables and not indicators, the principles employed are the same. Diagram 7.5 depicts a typical chart from their study. It shows the profile for Venezuela across the 18 variables used in their study as well as the main 'correspondence points' between all variables. These correspondence points are estimated by McGranahan *et al* across a moderately sized sample of nations and they identify the values on each variable corresponding to values on other variables. Once they have been able to establish these correspondences, the researchers can conclude: 'Venezuela shows several features quite the opposite of Ireland (also reported in the original study) – relatively low social indicators but high levels on certain economic indicators associated with oil (energy consumption, trade, per capita GNP). Owing to reliance on oil extraction, however, manufacturing is quite limited' (1972, 96). Statements like this can be very useful to identify, in a norm-referenced way, those variables or indicators on which a particular nation, or its education system, is doing relatively better than might be expected as well as those variables or indicators on which it is doing relatively worse. Indeed, the authors state: 'it may be noted that in practice, the profiles of 1960 would appear to suggest areas of possible action (eg lagging sectors) which in a number of cases governments have in fact pursued since 1960, presumably on the basis of other kinds of information' (1972, 97-8). Conceptually, therefore, this kind of presentation is very appealing – particularly if a hierarchical development view of education systems is adopted.

The graphing of values for a particular system on such a profile poses no problems. The establishment of correspondence points between particular values, however, is the crux of the entire analysis. The approach adopted by McGranahan *et al* is to prepare scatter diagrams for each pair of a core of four specially selected variables. A best fitting curve is estimated for these six curves by using the method of moving averages. Correspondence points are then estimated on three of the scales so that their values correspond to multiples of $100 or $200 on the per capita GNP variable – the fourth core variable – (although

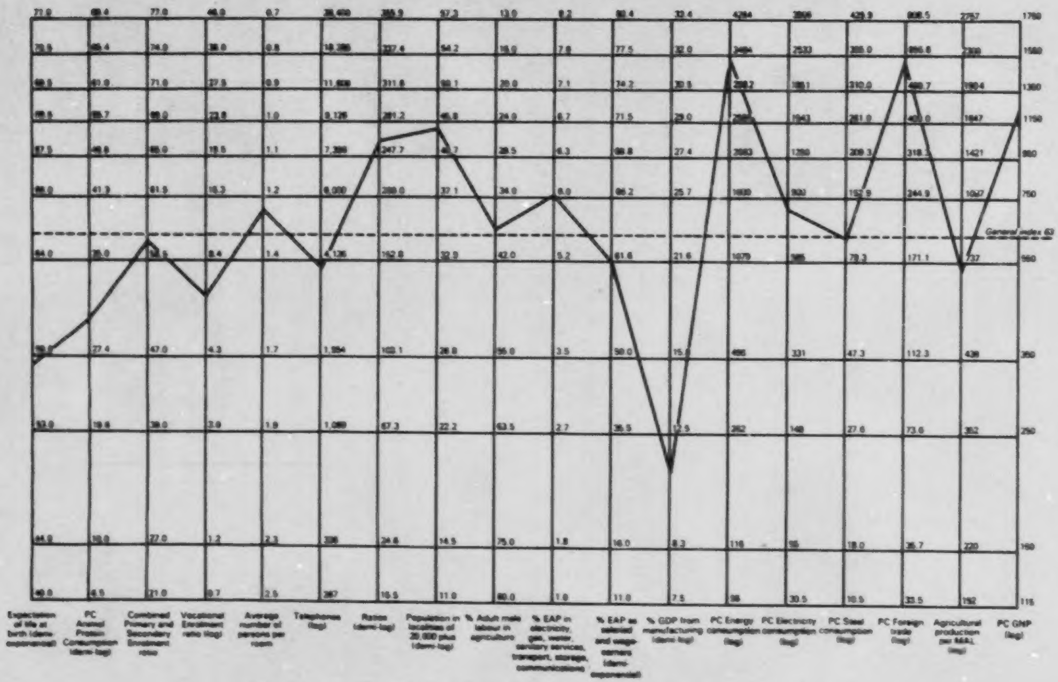


Diagram 7.5: *Socio-economic profile of Venezuela – 1960*
Source: McGranahan *et al* (1972, 20)

slight adjustments are subsequently made). The exact values allowing free exchange among the four variable values are finally determined from these 'best fitting lines' 'by a process of experimentation and rationalization' (1972, 86). Correspondence points are then estimated for the remaining 14 variables from the four core variables – again by the method of moving averages and 'a limited amount of reconciliation' (1972, 87).

The McGranahan *et al* approach does have some methodological flaws. For example, the correspondence points predicted by the variables do not necessarily have to coincide even approximately. This problem is magnified when the inter-relationships between pairs of variables being used approaches 0.0. Hence the estimates of equivalences can be very largely judgemental in many situations. In addition, imprecision of prediction is not accounted for and hence diagnosis of likely under- or over-achievement of an education system can be revealed as much by a weakness of the analysis as it can by a feature of the system being examined.

As a guide, this method of presenting indicators does have great intuitive appeal. Because of the need to have at least moderate inter-relationships among the indicators, the number of occasions on which it can be used are limited. Further research needs to be undertaken into the ways in which the correspondences can be better reported – perhaps even with the addition of bands indicating the standard errors of estimates.

7.3 Methods of presentation involving a number of indicators

The most complex and difficult situation in which to represent indicators graphically is encountered when values on a number of indicators have been estimated for each of a number of education systems. A matrix of values is thus available. One way to represent this matrix is to consider each indicator separately and graph the values for all systems. The simplest form of doing this has already been discussed in Section 7.1. Extensions of this concept are readily identifiable. Another way is to consider each system separately and to graph the values for each indicator. In this way, an individual profile of the characteristics of each system is developed. Such an approach facilitates a critical examination and evaluation of individual systems as well as comparisons of the similarities and differences shared by a number of systems.

One method of presenting a number of indicators for one system is used in Diagrams 10.2 and 10.4. There a slight modification is introduced in that, rather than draw the profile for each system, the average profile is presented for a number of systems identified analytically as being similar. The profiles in these two diagrams demonstrate the usefulness of this method of presentation. Characteristics on which systems are relatively high or low can be readily identified – very much more so to most people than tabular reports of statistical figures. The similarities in profile shapes are also easier to detect as too is the extent to which indicator values make two or more systems increasingly dissimilar. For example, in Diagram 10.2, clusters 1, 3 and 7 contain countries with relatively similar profile shapes (or directions). The position of the profiles on the vertical axes (ie the heights of the columns) is therefore the principal feature distinguishing among these three clusters.

Two principles should be adhered to when presenting profiles for systems across a number of indicators. One is that the values for each indicator should be standardized in some way. This does not necessarily imply that standard scores[1] have to be calculated for each indicator. Where it is possible, however, comparable scales across indicators should be used. Where indicators using varying statistical bases are to be represented simultaneously in the one profile in terms of their original values (perhaps to allow criterion-referenced comparisons rather than norm-referenced comparisons), those indicators using the same kind of statistical basis should be grouped together. Appropriate scaling procedures can then be applied to ensure comparability of heights etc as is necessary.

A second principle for drawing profiles is that high values on each indicator should be graphed in the same direction regardless of the implications of the feature being measured by the indicator. A consistent policy on representing high (positive) values in this way can avoid considerable confusion in interpreting a profile. It could be, for example, that a high value on one indicator should ideally accompany a low value on another. Attempts to show a particular system conforms to this pattern by drawing a relatively straight line for its profile are generally not clear. Indeed they can easily confuse the interpretation made by observers less familiar with a given situation being depicted.

The concept embodied in this principle emphasizes that no normative interpretation of 'good' or 'bad' can be imposed on the shape of a profile *per se*. Instead such a value must be inferred through the values depicted by the profile. In other words, a profile showing high values across all indicators in a defined set does not necessarily reveal that a system being represented is a 'good' system. Perhaps on some of the indicators, high values reveal features on which planning aspirations have not been realized. Instead of attempting to adopt such an approach, an 'ideal' profile can be developed with the high points and the low points included as appropriate. This would then form a basis with which to compare the profile for any particular system.

---

[1] The profiles in Diagrams 10.2 and 10.4 have been standardized to allow comparisons among the indicators in each diagram.

An alternative to this mode of presentation is based upon the recommendations of Chernoff (1973). It assumes we are more used to looking at people and noting their individual differences than we are to looking at abstract statistical graphs. We can, for example, readily note and quickly evaluate a number of personal characteristics. This is really exactly the same task which must be undertaken when confronted with a number of profiles. Any method incorporating a recognition technique must therefore develop a *gestalt* of the characteristics of an education system in a readily identifiable manner. The way described here develops this *gestalt* by ascribing particular indicators to particular features of a face.

The data used to demonstrate this technique are reported in Table B.5. They are factor scores calculated for 11 of the national education systems analysed and they relate to the ten process factors identified in the factor analysis described there. Any indicator values could have been used, however. The education systems are chosen after a trial typology has been formed in the way to be described in Chapter 10. They are thus systems which have been demonstrated empirically as being different from one another. The score for each factor is assigned to a particular facial feature. According to the value for a given education system, the facial feature is then changed. The results[1] are reported in Diagram 7.6. The female involvement factor, for example, has been assigned to the characteristic of 'smile on mouth'. The higher the values for this indicator for an education system, the bigger the smile. Hence Australia is shown to have (proportionately) more female pupils and teachers than Iran and hence the 'Australian face' has a bigger smile than the 'Iranian face'. Openness of eye reflects a high proportion of second-level students in the general education stream. Hence the USA and Iran are much more 'open-eyed' than Austria, Israel or the USSR which all have a high concentration (comparatively) on vocational education at the second-level. All ten empirically defined factors have been assigned to a particular characteristic and these assignments are listed in Table 7.2. All faces should be compared against the average profile – represented by 'average shape' – which results when the factor scores for all ten indicators equal zero.

7.4 Résumé

The development of methods for presenting indicators to consumers is a very necessary and integral part of overall indicator development. Recent indicator documents such as *Social Indicators* have tried to report properly both indicator values and variable values. Such attempts

---

[1] The assistance of Dr Michael Bailey of the School of Education, Macquarie University in writing the computer program for the diagram is gratefully acknowledged.

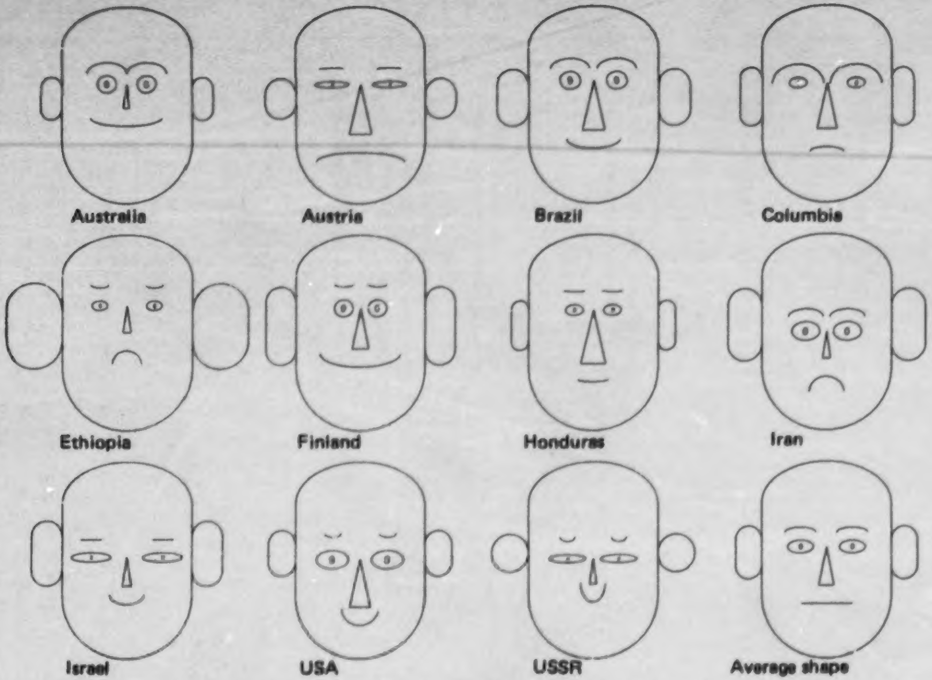


Diagram 7.6: *Representation of empirically defined education system characteristics – selected nations, 1960*

Facial characteristic	Empirically defined indicator number	Interpretation
Width of mouth	1	The wider the mouth, the more extensive is basic educational development.
Smile on mouth	2	The greater the smile, the more extensive is female involvement in educational provision.
Length of ears	3	The longer the ears, the greater the development emphasis on second-level education.
Width of ear	5	The wider the ear, the greater the development emphasis on providing an equality of involvement for female pupils.
Overall size of nose	4	The larger the nose, the greater the proportion of third-level enrolments in law/social science courses. The smaller the nose, the greater the proportion in humanities/education/fine arts courses.
Curvature of eyebrow	7	The greater the upward curvature of the eyebrows, the greater is the emphasis on expanding teacher education facilities.
Openness of eye	6	The more open the eyes, the greater the proportion of second-level enrolments in the general education stream.
Extent of sloe-eyedness	8	The more sloe-eyed, the greater the relative extent of pre-school provision.
Distance between pupils	10	The further apart the pupils, the greater the proportion of third-level students in natural science courses.
Position of eye	9	The lower the position of the eyes, the higher is the general quality of educational provision.

Table 7.2: *Guide to the interpretation of Diagram 7.6*

as this are few and far between, however. There is a variety of ways in which indicators can be presented and these emanate from at least three distinct purposes for and kinds of indicator presentation. The extent of the variety of possible methods is indeed wide – and the use of human facial features testifies to this. Although only some of the possible methods which use printed means of communication have been discussed here, non-printed methods using TV display screens, audio-visual productions, etc are also claimed by their proponents to have great potential. Further work in this area of communication is needed, however. It is becoming increasingly important to undertake discussions with diverse groups of people about the implications of educational trends in particular but also of social trends in general which are measured by defined sets of indicators.

Chapter 8

The Use of Indicators in Educational Planning

The basic sets of indicators proposed in Chapters 5 and 6 for national and regional education systems form a basis from which people engaged in particular planning situations must extrapolate. In some situations, data may be available and the education system objectives might be such as to allow all indicators as proposed to be adopted for use. In other circumstances, a selection has to be made from the lists as recommended. In addition, other indicators will sometimes have to be defined to measure system objectives not covered by the indicators proposed here. The unavailability of data can also cause modifications to be made to the basic sets as well as to the definitions of the recommended indicators.

To demonstrate how some of the indicators function in practical planning situations, two contrasting examples are now discussed. In the first of these, a number of input and process indicators are selected to help evaluate progress in a national education system in relation to certain stated objectives. The country chosen for this discussion is the Kingdom of Saudi Arabia. In the second example, nine process indicators are selected to demonstrate the way in which regional education systems development can be monitored. Here the country chosen is Indonesia. In both examples, modifications to the basic definitions of some of the indicators are required. No example is provided in this chapter to demonstrate the use of output indicators or of education system-societal congruence indicators. The application of such indicators parallels the procedures to be followed here. Hence their omission should not be interpreted as implying possibly diminished importance or usefulness of these types of indicators nor that there are special difficulties involved in their calculation. Neither type could be naturally incorporated into the two situations chosen.

8.1 The use of indicators to help evaluate planned national system development

The focus for the example of evaluating planned national education system development is that part of the Saudi Arabian First Development

Plan 1390-1394 AH (1970-1974AD) concerned with the education sector. It is therefore not a complete educational plan in its own right but it is the type of approach which is frequently met in many planning situations.

The statements made about education in the planning documents are of two kinds. Some statements describe specific objectives which focus on very particular outcomes intended by the development plan. Some of these objectives set a number of targets in terms of gross enrolments. For example:

> to increase present total enrolment (for first-level boys) by 55 per cent by the completion of the Plan. (p98)

Statements such as this do not need indicators defined for their evaluation. They are concerned with special aspects of the system and consequently a variable is the more appropriate quantity to define. Other specific objectives do however lead to an evaluation by means of indicators. Examples of these objectives are presented later in this section.

Also included in the planning document are statements which describe general objectives. These intended outcomes establish the framework for the education sector part of the development plan. They reveal what is intended for the overall development of the Saudi education system. An example of these general objectives is:

> The Development Plan for education looks to continued expansion of opportunities for education at all levels from elementary through college. (p24)

Both specific and general objectives are required in any development plan and both kinds must be referenced in an evaluation. In the present context, attention is focused especially on those objectives which can be evaluated by means of indicators.

The above example of a general objective provides an important starting point for an evaluation for it identifies an intention on the part of the Saudi planners to expand their education system quite significantly. Classification of the objective must be ambiguous. In one sense it can be regarded as commenting on the levels of input into the system and hence it is capable of being evaluated through the 'involvement of human resources' input indicator. In another sense, the objective can be classified into the process sub-division and therefore evaluated through the 'relative availability of student places' indicator. In any comprehensive evaluation of this objective as stated, both indicators should be used.

Involvement of human resources

Table 8.1 reports the index numbers for the 'involvement of human resources' indicator together with the original data for the component variables. Values refer to the years. Unfortunately data for the last year of the planning period are only available for ministry of education schools (boys only) and hence the data for the previous year must be taken instead.[1] They allow most of the planning period to be covered, however. The data for 1965 are included to establish a clearer trend in the indicator values as well as to reveal what happened prior to the commencement of the development plan.

Variable	Year 1965	 1970	 1973
First-level enrolment ratio (males)	22	32	41
First-level enrolment ratio (females)	8	16	21
Second-level enrolment ratio (males)	3	11	13
Second-level enrolment ratio (females)	1	3	5
Third-level students per 10,000 population	5.4	11.0	17.6
Teachers per 10,000 population	19.1	30.5	44.9
Indicator value	49	100	142

Table 8.1: *'Involvement of human resources' input indicator – values and basic data, Saudi Arabia, 1965-73*

A significant increase in the involvement of human resources in the Saudi education system is demonstrated by the indicator values over both the 1965-70 period and the 1970-73 period. Against the base of 100 set for 1970, the 1965 value of 49 shows that a 15.3 per cent annual rate of increase occurred up until 1970. From 1970 to 1973, the rate of increase slowed a little to 12.4 per cent – still quite a substantial rate of increase. The higher rate calculated for the earlier period is attributable to the very small values for the component variables (shown in Table 8.1) for 1965. The indicator values show many more people are engaged in the Saudi Arabian education system and the increase in their number is far greater than that which might have been expected under normal pro-rata population increases.

Availability of student places

The process indicator measuring the relative availability of student places complements the previous input indicator in the comment made about progress towards the attainment of the stated objective. The values for this indicator are reported in Table 8.2 together with the data for each of the component variables. A substantial increase in the indicator value is shown to have occurred during both periods being considered. Since 1966, more students have remained in the Saudi Arabian education system for longer and hence the total educational pyramid in Saudi Arabia has substantially enlarged. The average annual

---

1. All basic data used in this section were published by the Saudi Arabian ministry of education in their *Education Statistics* for 1390/91 and 1393/4.

241

rate of increase in the value for this indicator for the 1970-73 period (6.7 per cent) is slightly greater than that for the 1966-70 period (5.1 per cent). This higher rate could be attributable, in part at least, to efforts by administrators and planners to encourage increased participation by the population in all levels of the education system in accordance with the objectives of the development plan. This process indicator, in conjunction with the input indicator discussed above, reflects that 'expansion of opportunities for education at all levels from elementary to college' (as stated in the objectives cited) does seem to have continued.

System levels to be ratioed	Sex	Percentage ratios for year 1966	1970	1973
Pre-school to first-level	Male	1.08	1.24*	1.51
Pre-school to first-level	Female		1.59*	1.68
Special education to first-level	Male	0.30	0.40	0.40
Special education to first-level	Female	0.12	0.80	0.11
Second-level to first-level	Male	18.10	24.69	27.99
Second-level to first-level	Female	6.43	13.23	20.89
Third-level to second-level	Male	9.62	10.88	12.15
Third-level to second-level	Female	2.45	3.95	4.72
Indicator value		82	100	121

* The combined value for the pre-school to first-level enrolment for 1970 to be divided into the 1966 value is 1.35.

Table 8.2: *'Availability of student places' process indicator – values and basic data, Saudi Arabia, 1966-73*

The values for this process indicator also reveal that two other objectives stated in the plan could be claimed to have been substantially achieved during the 1970-73 part of the planning period. They are, for boys' education:

> to provide facilities for all elementary school graduates who would be expected to apply for intermediate school enrolment; (p98)

and for girls' education:

> to improve the enrolment relationship between intermediate level education and elementary education from the present 9.3 per cent (intermediate enrolment as a percentage of elementary enrolment) to 17.4 per cent at the completion of the Plan. (p99)

Both of these objectives require data on specific variables for their evaluation. The data reported in Table 8.2 are not exactly what are required. If, however, the total second-level to elementary enrolment percentage is accepted as being a surrogate of the specific intermediate level to elementary level percentage, a comment can be made. The ratio for boys rose from 25 per cent in 1970 to 28 per cent in 1973 while for girls it increased from 13 per cent in 1970 to 21 per cent in 1973. These are significant rises – especially considering the short period being considered. Both additional specific objectives can therefore be said to have been achieved. The process indicator as discussed above puts these specific increases into the total perspective and thus comments more globally, and, in some ways, more significantly about the overall expansion of student places in the Saudi education system.

Female involvement

No single objective in the development plan relates directly to a policy of increasing the overall involvement of females in the Saudi education system. A number of statements do however point to specific aspects of such a policy. Two examples demonstrate this claim:

> to increase present [1389-90 AH] elementary school enrolment [for girls] by 95 per cent by the completion of the plan; (p99)
>
> to increase general secondary enrolment [for girls] from the present level of 350 students to over 4900 students by the completion of the Plan. (p99)

Both of these statements, together with other statements for other levels of the system, indicate that the Saudi planners and administrators are embarking on a program of significantly increasing the involvement of females in education, particularly amongst pupils.

Values for the female involvement indicator together with values for its component variables are reported in Table 8.3. A most substantial increase is shown to have occurred in the inclusion of women in both the student body and the teaching force. During the 1966-70 period, the average annual rate of increase was 15.1 per cent. This rate then slowed to 11.9 per cent during 1970-73. Nevertheless both rates of increase reflect a radical change in the composition of the Saudi education system. Such change had to occur if the demands of the Saudi development projects were to be met by Saudi nationals rather than by expatriates. The magnitude of the change as shown here should be heartening to those closely connected with the Saudi Arabian education system.

Financial involvement

Given the significant changes which have occurred and which are revealed by the indicators so far presented, it is useful to measure the change in the level of financial involvement over the period. This can be effected through the appropriate input indicator– number 3 in Table 5.2. The values for this financial involvement indicator and for its component variables are reported in Table 8.4.

The values reveal an unexpected pattern in view of the expansion of the system which has taken place. Quite a substantial decrease occurred over the period. During the 1965-70 period, the decrease averaged

Level and type	Females as percentage of total in year		
	1966	1970	1973
First-level pupils	25.0	31.3	34.2
Second-level pupils	11.6	18.0	27.9
Third-level pupils	3.0	8.9	13.1
First-level teachers	0.0	27.1	29.0
Second-level teachers	7.4	13.6	21.9
Third-level teachers	0.0	6.9	11.0
Indicator value	57*	100	140

* The value of 57 excludes the two zero values. If these are included in an arithmetic average, the 1966 value decreases to 38.

Table 8.3: *'Female involvement in the provision of education' process indicator – values and basic data, Saudi Arabia, 1966-73*

Variable	Year		
	1965	1970	1973
% GDP spent on education	6.3	6.7	5.8
% total government expenditure spent on education	13.0	10.0	8.2
Indicator value	112	100	85

Table 8.4: *'Financial involvement' input indicator – values and basic data, Saudi Arabia, 1965-73*

about 2.2 per cent per year. This rate accelerated to an average of 5.3 per cent over the 1970-73 period. In normal circumstances, this decrease might indicate an alarming situation and one that would signal a significant decline in conditions within the system. In the Saudi Arabian situation, however, the decline is due to the enormous growth rates brought about by the revised oil price agreements from about 1969. The staggering growth rates manifested by the Saudi economy are well known. Over the three years to 1970, the gross domestic product rose by 35 per cent while oil incomes rose by 48 per cent. Over the next three years to 1973, GDP rose by a further 59 per cent while the oil sector of the economy rose by 84 per cent. The decrease in the financial involvement indicator as noted here therefore is due to the distribution of the extra finances to other sectors and government projects as well as to the inability of the education system to accommodate such a rapidly rising income. As no comments are made within the development plan to indicate the patterns educational expenditures might have followed, no objective can be evaluated in light of this particular indicator.

Orientation to second-level studies

Among the general objectives of the plan, it is stated that:

> The Plan provides for more diversification in educational offerings once the student has completed his basic general education. Industrial education for boys at the secondary level is being expanded. (p24)

In the specific objectives, the statements concerned with technical education all imply an emphasis on expanding this form of education. Merely reporting the percentage of second-level students enrolled in vocational education courses does not however assess the relative 'diversity' or concentration on this particular form of education. The C index of concentration combines the percentage of senior secondary students in the general education and the teacher training strands and does assess the diversity. It takes into account the relative percentages in these two strands and assesses their concentration relative to that in the vocational education strand. Hence all three strands of education at the senior secondary level are either directly or indirectly accounted for. The index required here was defined in Table 5.4 as 'the orientation of second-level studies'. Values for the indicator and its component variables are reported in Table 8.5.

Sex	**Second-level strand**	**% of total enrolment in year**		
		1966	**1970**	**1973**
Males	**General education**	**55.8**	**65.3**	**67.0**
	Teacher training	**30.2**	**30.8**	**26.8**
	Indicator value	**0.40**	**0.52**	**0.60**
Females	**General education**	**17.9**	**21.0**	**54.6**
	Teacher training	**82.1**	**79.0**	**42.0**
	Indicator value	**0.71**	**0.67**	**0.47**

Table 8.5: *'Orientation of second-level studies' process indicator – values and basic data, Saudi Arabia, 1966-73*

For males, there has been an increasing concentration in the orientation within their senior second-level studies. The value of C rose steadily from 0.40 for 1966 to 0.60 for 1973. This increase is attributable to the increased percentage of students enrolling in the general education strand and to the corresponding decline in the percentage of students enrolled in the teacher education strand. The concentration of male second-level students in general education courses has therefore increased especially at the expense of vocational and technical education.

For females the values for the C index of concentration decreased over the 1966-73 period. They thus provide a marked contrast to the pattern established for males. The decrease changed the C value from 0.71 to 0.47. An overall dispersion in the education offerings at the senior secondary level thus occurred for girls. The underlying pattern of the variables is a confused one, however, and quite atypical. What happened is that general education courses increased in the percentage of students they attracted from a low 18 per cent to 55 per cent while the percentage of students in teacher training decreased from a very high 82 per cent to 42 per cent. Enrolments in the vocational and technical courses thus accounted for no more than about 3 per cent of total enrolments at this level. The more equal balance between general education and teacher training enrolments thus implies that there is no concentration of female enrolments in a particular strand in 1973. The decreased value for the C index reflects the greater balance.

This result is certainly not an intended outcome according to the statements in the development plan. The above objective, for example, stated that emphasis was to be placed on industrial education for boys. This has not happened. Elsewhere the plan also states an objective:

> to increase the emphasis on teacher training at the second-level. (p98)

This has not been achieved. Indeed a decrease has occurred instead. Offsetting both of these decreases has been the increasing concentration on general education. If trends elsewhere can also be taken as a guide, the Saudi pattern tracing development over the 1966-73 period will continue during the next planning period also with the consequence that a further increase in the concentration of students in the general education strand will occur.

Orientation to third-level studies

No definite indication as to the intention of the Saudi Arabian planners can be ascertained from the development plan with respect to the intended or desired trends in the indicator measuring the orientation to third-level studies. Comments are made in both the general and the specific objectives as to particular increases planned for the enrolment numbers and/or in particular third-level institutions. For example, one objective, stated for the University of Riyadh, is

> to strengthen all present eight faculties through the addition of staff and facilities as necessary to meet enrolment increases. (p99)

For the College of Petroleum and Minerals, it is stated:

> to expand undergraduate enrolment to teach a total of 1150 students by the end of the plan period. (p99)

For the College of Education at Mecca, it is stated:

> to increase enrolment to approximately 950 students by the completion of the plan period. (p99)

For the King Abdul Aziz University, it is stated:

> to expand the Faculty of Arts and Human Sciences from one to three departments by 1391-92. (p99)

> To establish a Department of Accounting in the Faculty of Economics and Administration in 1392-94. (p99)

These objectives plus some of the general statements made in the development plan indicate a policy of expanding the humanities, economics and education faculties in particular as well as the natural science and engineering faculties.

Estimates across all third-level faculties for males and for all three faculties available to females result in the indicator values reported in Table 8.6. Also reported there are the values for each of the component variables. For males, the C index decreases in value from its 1966 level of 0.48 to 0.25 for 1973. This reveals that third-level education in Saudi Arabia has become much less concentrated across the seven faculties. The values for the component variables confirm this finding. The percentage calculated for the humanities/law enrolments decreases while all other six percentages increase. The extreme dominance of the humanities/law combination in third-level offerings for males has therefore been significantly reduced and thus the diversity in enrolments among the faculties can be said to have increased. If this pattern had been duplicated in the University of Riyadh also, then the faculties on offer there might have all expanded but not by equal amounts. A direct evaluation of that particular objective is not possible (nor appropriate) here.

Sex	Faculty	% of total enrolment in year		
		1966	1970	1973
Males	Humanities and law	66.9	52.9	41.8
	Education	4.5	9.3	13.0
	Social studies	14.5	13.9	15.8
	Natural science	4.5	7.9	8.2
	Engineering	7.3	11.7	15.2
	Medicine	0.8	2.3	3.3
	Agriculture	1.7	2.1	2.9
	Indicator value	0.48	0.33	0.25
	Relative concentration between humanities/law/education faculties and natural science/engineering faculties	61.13	15.44	6.42
Females	Humanities and law	76.3	43.6	47.1
	Education	0.0	30.4	35.5
	Social science	23.7	26.1	17.4
	Indicator value	0.64	0.35	0.38

Table 8.6: *'Orientation of third-level studies' process indicator — values and basic data, Saudi Arabia, 1966-73*

The relative concentration of enrolments across faculties especially as implied by the last four objectives listed above, can also be assessed by ratioing particular C indices. One such index is based on only the humanities/law/education faculties while the other is based on only the natural science/engineering faculties. This ratio, reported in Table 8.6, decreases over the period being examined from 61.13 in 1966 to 6.42 in 1973. This is a very significant decrease indeed. It is caused especially by the trend, as noted above, of a very large decrease in the percentage of enrolments in the humanities/law faculties.

Females in Saudi Arabia are enrolled in three faculties at most during the period being considered here. The C index calculated over these faculties shows a very significant decrease in value over the first part of the period and then a very slight increase in value over the latter part. The decrease is attributable to the emphasis on education courses which were not available in 1966 but which attracted over one third of the total 1973 third-level female enrolments. Although it is not possible to evaluate this trend *vis-à-vis* a particular stated objective in the development plan, it would be difficult to argue that the pattern identified here would be counter to that desired by the Saudi planners. An expected outcome from it would be the graduation of Saudi women in a variety of fields. These graduates could then take their place in the work-force – perhaps even as replacements for some of the expatriate short-term workers.

Internal efficiency

A general objective of the development plan states an intention to undertake 'efforts concentrated on measures that will improve efficiency [in the education system]' (p24). Efficiency in this sense can be interpreted as being the flow through the education system. At the first-level, the extent to which this objective has been attained can be evaluated by the process indicator measuring the expected number of years of education received by an entrant as well as by the process indicator measuring the expected number of grades attained. Values for both of these indicators, as well as the transition matrices from which the values were calculated, were presented and discussed in Section 3.3. The number of years of education expected to be received by an entrant reduced from 6.94 to 6.68 over the 1969-73 period. The average number of grades completed rose from 4.75 to 5.18. Both of these changes show an improved rate of flow of students through the Saudi first-level. Hence, to the extent measured by these two indicators at least, the general objective relating to improved efficiency, as stated in the development plan, can be said to have been achieved.

Although the Saudi Arabian development plan does state other objectives for the education sector, these are not appropriate to discuss in the present context. They are related to specific features such as the establishment of targets for pupil-teacher ratios and to attaining particular intake levels for individual institutions. Such objectives need to be evaluated using specifically defined variables.

Summary

This presentation – albeit limited in scope and briefly presented – demonstrates that it is possible, beneficial and efficient to evaluate a plan for education system development using a set of indicators. Some of the stated objectives are shown to have been very satisfactorily attained while other objectives are shown not to have been attained. Where it was possible to use indicators, a large amount of data, covering some 47 variables for each year, has been condensed to form eight indicators. This reduction in no way however affects the ease with which these particular aspects of the education system should be evaluated. Indeed it improves the ability of planners, administrators, researchers and evaluators to form a *gestalt* of the overall pattern of Saudi development.

8.2 The use of indicators to monitor regional system development

One of the major planning uses predicated for indicators is the monitoring of education system development. This can be effected either on a cross-national basis or on an intra-national basis. To demonstrate how education systems located within particular regions of a single country can be monitored as well as to demonstrate how some of the process indices in particular comment on development in a specific situation, an example is presented for the country of Indonesia. This example contrasts with that for Saudi Arabia in the last section in that here no evaluation is conducted with respect to a development plan. Certainly, for Indonesia, this would have been possible, for in the period to be considered, the Second Five Year Plan – Repelita II – was being implemented. Instead, in this section, a situation common to that in many countries which have no formal plan is simulated so that a monitoring rather than an evaluation is demonstrated.

The years for which indicator values are calculated are 1974 and 1976. The reasons motivating this choice are related to data availability and reliability and not to any features particular to those years. Ideally a longer time between observations would have been desirable, but this was not possible. Over the two year period, therefore, only initial and tentative trends among the provincial system development patterns can emerge.

Data availability allows nine of the process indicators listed in Table 5.4 to be estimated for each province.[1] Values for a sample of ten Indonesian provinces[1] are presented in Table 8.7 for eight of these indicators. Four of the indicators reported there (numbered 1-3 and 7) use index numbers calculated with respect to a base of 100 which is set equal to the national average for 1974. Hence the values for these indicators can be interpreted both across provinces within each year as well as for each province across the two years. A further two indicators (numbered 8 and 12) are C indices of concentration.

---

[1] A fuller discussion of these indicators plus two input and two output indicators is presented in Johnstone (1978a).

Table 8.8 reports the other indicator – based on the Gini coefficient – calculated across the country as a whole. The relative location through the Indonesian archipelago of the ten provinces used in this demonstration is shown in Diagram 7.2.

Availability of student places, schools and classes

The first three indicators have to be defined in a slightly different way to that described in Table 5.4 so that they can better suit the Indonesian context. Here pre-school, third-level and special education are excluded. Two variables only are combined to form the indicator value. These are:

☐ the ratio of the characteristic at the junior secondary level to the first-level

☐ the ratio of the characteristic at the senior secondary level to the junior secondary level.

The three indicators therefore comment respectively on the relative availability of student places, schools and classes for first- and second-level education.

All three indicators are closely inter-related and they comment in very similar ways on the levels of educational provision in the ten provinces sampled here. The highly urbanized and special areas of Jakarta and Yogyakarta have indicator values well above those for all the other provinces. The relative rankings of the provinces across the two years are basically the same although different growth rates can also be detected emerging. East Nusatenggara, for example, shows definite patterns of increase which are quite substantial. Yogyakarta

---

[1] The 26 Indonesian provinces referred to in this chapter and whose geographical locations are depicted in Diagram 7.2 are very diverse in their levels of economic and social development. The five provinces on the island of Java are amongst the most developed partly because of the presence of the three largest cities in Indonesia – Jakarta and Yogyakarta (which both form separate provinces) and Surabaya in East Java. In the island of Sumatra, the province of West Sumatra is also well developed. The remaining provinces in that island are at a lesser development level with Lampung being least developed. The provinces in Kalimantan, Sulawesi and East Indonesia tend to be in the lower half of the Indonesian development scale with Irian Jaya, the Nusatenggaras and Central Kalimantan towards the bottom. For these last provinces in particular, there are extreme difficulties in communicating with Jakarta. Bali, although in the East Indonesia group is fairly well developed mainly because of the influence of the heavy tourist industry.

Province	Process Indicator 1. Student places 1974	1. Student places 1976	2. Schools 1974	2. Schools 1976	3. Classes 1974	3. Classes 1976	7. Female involvement 1974	7. Female involvement 1976	8. Orientation second-level 1974	8. Orientation second-level 1976	10. Years of education 1974	11. Expected grade 1974	12. Qualified teachers 1974	12. Qualified teachers 1976
(a) Java														
Jakarta	125	131	124	130	131	136	113	116	0.28	0.31	NA	NA	0.31	0.39
West Java	98	100	97	97	98	101	101	103	0.24	0.24	5.45	4.56	0.46	0.48
Central Java	102	103	97	95	102	102	94	99	0.15	0.12	5.81	4.76	0.43	0.65
Yogyakarta	127	131	123	118	125	126	101	104	0.20	0.20	7.24	NA	0.51	0.65
East Java	99	100	98	96	100	100	100	105	0.17	0.16	6.17	5.20	0.41	0.63
(b) Eastern Indonesia														
Bali	105	104	102	97	105	103	86	92	0.17	0.20	5.75	4.84	0.61	0.73
W. Nusateng.	97	98	96	98	97	98	78	85	0.21	0.25	4.63	3.44	0.37	0.59
E. Nusateng.	91	98	90	94	90	95	84	88	0.19	0.24	5.99	4.49	0.38	0.62
Maluku	101	104	97	95	100	98	101	103	0.20	0.22	6.28	4.68	0.27	0.44
Irian Jaya	86	92	88	89	88	92	84	89	0.19	0.17	5.61	4.22	0.27	0.27
Indonesia	100	103	100	99	100	103	100	105	0.21	0.21	NA	NA	0.38	0.53

Notes:

1. Numbers in front of the indicator names refer to the numbers of the process indicators as defined in Table 5.4.
2. Indicators 1–3 and 7 are reported as index numbers each standardized to a value of 100 set equal to the national 1974 average.
   Indicators 8 and 12 are computed as C indices of concentration.
   Indicators 10 and 11 use the usual formulae as described in Chapter 3.
3. NA = Insufficient reliable data were available to allow the computation.

Table 8.7: *Summary of values for selected process indicators monitoring the development of education systems in selected provinces in Indonesia – 1974 and 1976*

increases significantly on only one indicator while it regresses quite significantly on another.

The 'availability of schools' indicator (number 2) decreases in its national value by one point. This reflects perhaps that a building programme was in operation between 1974 and 1976 which has effected a very slight (and probably insignificant) decrease in the relative availability of schools which allow students to move from one level or stage of the education system to the next. In the short term, a decrease in the indicator value probably only reflects different rates of implementing a building program. A longer term decrease – especially one which equalled a number of points – would be quite serious. It would probably reflect a short-sighted policy in which a demand for education at one level of the system was being fulfilled by the provision of schools at that level but, by so doing, equal availability of schools for graduates of that level was being denied.

Female involvement

The female involvement index shows similar overall patterns to those revealed by the first three indicators. All provinces increase their values over the two years – thus showing increasing participation rates of females in the education systems in these Indonesian provinces. Rates of increase for the East Indonesian provinces exceed those for the Javanese provinces although overall their involvement indicator values still lag behind those existing for Java. Maluku is the exception to this observation in that it is more like the Javanese provinces than it is like its geographically closer neighbours. Another noticeable difference between the values for this indicator and those for the first three indicators is that the provinces are here ranked in a very different order. This alternative ranking reflects the differences between the characteristics being measured and thus confirms the claim that it is necessary to have a number of indicators to monitor system development.

Orientation of second-level studies

The indicators measuring the orientation of second-level studies and the availability of qualified teachers at the first-level (numbers 8 and 12) are both based on the C index of concentration. The lower the indicator values, the less concentrated is the characteristic being measured in the education system. The values for the orientation of second-level studies (at the upper secondary level only in this case) are based on the general education and the teacher training strands. The low values for this indicator thus reveal a lack of concentration on one type of education. Indeed the basic data (not reported here) reveal about 50 per cent of students in both Jakarta and West Java are enrolled in general education courses while about 45 per cent are enrolled in technical or vocational education courses. In Central Java, which has the lowest of all ten values reported, the proportions are about 38 per cent and 55 per cent respectively. In none of these provinces therefore is there a tendency to stress one type of education significantly more than another. Little trend over the two year period is discernible except for Jakarta which increases its degree of concentration (on general education) and Irian Jaya which decreases its concentration in these two areas (and hence increases its concentration on technical and vocational education).

Availability of qualified teachers

For the indicator measuring the availability of qualified teachers at the first-level (number 12) the C values are moderately high. They have been computed using the percentage of teachers in the 'unqualified' and 'qualified' categories. 'Semi-qualified' teachers have thus been excluded from the computation of the C values although they have been included in the total number of teachers for the computation of the two required percentages.

Considerable variation is in evidence among the provinces. Somewhat surprising is that the capital city province – Jakarta – has nearly the lowest 1974 C value. Bali has the highest value followed by Yogyakarta. The value for Bali indeed reveals a very significant concentration of teachers in the 'qualified' category. The C index value demonstrates this concentration better than merely citing the variable measuring the percentage of the teaching force who are qualified. It has this advantage because, by including the component of the 'unqualified' category as well, it identifies the distribution of qualifications in a province and not just the proportion in one category. For example, the value for Bali is considerably above that for Jakarta. Hence there are many more semi-qualified teachers in Jakarta than there are in Bali. The C value shows that it is not the percentage of unqualified teachers which is high in Jakarta. The basic 1974 data are:

- ☐ for Bali, 78 per cent of teachers qualified with 21 per cent semi-qualified
- ☐ for Jakarta, 55 per cent qualified and 40 per cent semi-qualified.

These confirm the interpretation of the C index made above.

The C values for this indicator for most of the ten provinces reported in Table 8.7 increase significantly over the short period. This is especially so for East and West Nusatenggara and for Central and East Java. These increases reveal shifts from the 'semi-qualified' to the 'qualified' category. They reflect the effects of the major teacher upgrading programs held during the period being considered.

Expected years and grades of education

In Table 8.7, indicators numbered 10 and 11 measure the expected number of years an entrant to the first-level can expect to spend in that level and how many grades that entrant can expect to complete. For all

Indonesian provinces, the formal length of the first-level is six years. Hence the values in the table should be compared to this figure.

Considerable variation exists both amongst the values for each indicator and between the two indicator values for each province. In West Java, for example, a child entering grade 1 of the first-level could expect to spend about five and a half years in that level and to complete about four and a half grades of that education. In Maluku, however, a child could expect to spend more than six years in the system although the grade reached at the end of that time would be equivalent to just a few months more than that received by the West Java child. The Java provinces and Bali appear, on these figures, to offer (at least in 1974) a more efficient flow of students through the first-level than the other East Indonesian provinces.

Distribution of qualified teachers

Values for one additional process indicator are reported in Table 8.8 for both 1974 and 1976. This indicator (number 13 in Table 5.4) measures the equality in the distribution of qualified teachers across all 26 Indonesian provinces and not just the ten provinces listed in Table 8.7. If there is an equal distribution, then students at a particular level are given equal opportunity to be taught by teachers qualified to teach at that level irrespective of the province in which they live.

Level	1974	1976
First-level	0.035	0.041
Junior secondary level	0.097	0.093
Senior secondary level	0.185	0.162

Table 8.8: *Indicator values measuring the equality of the distribution of qualified teachers across Indonesian provinces – 1974 and 1976*

The indicator is based on the Gini coefficient. The calculation of the value for each level of the system is accomplished by inter-relating, through Equation 3.9, the percentage of all qualified teachers in Indonesia who teach in each province and the percentage of total enrolments at that level who attend school in each province.

The values for each level for the two years, as reported in Table 8.8, show that there is considerable equality in the distribution across provinces. All six values are relatively small. At both the first-level and the junior secondary level, the values are not at all significant and thus the Lorenz curves would deviate from the lines of equality only marginally. At the senior secondary level, the coefficients are slightly higher. There is therefore a slight inequality in the two distributions at this level although even this has decreased over the two year period.

Diagram 8.1 demonstrates the nature of the slight 1976 inequality at the senior secondary level. The Lorenz curve drawn shows that provinces such as Jambi, Central Kalimantan and Riau have fewer qualified senior secondary teachers than they might expect, given their enrolments, while all the provinces in Java plus North Sumatra have relatively more qualified teachers. Overall, however, there is considerable equality in the distribution of qualified teachers throughout the Indonesian provinces.

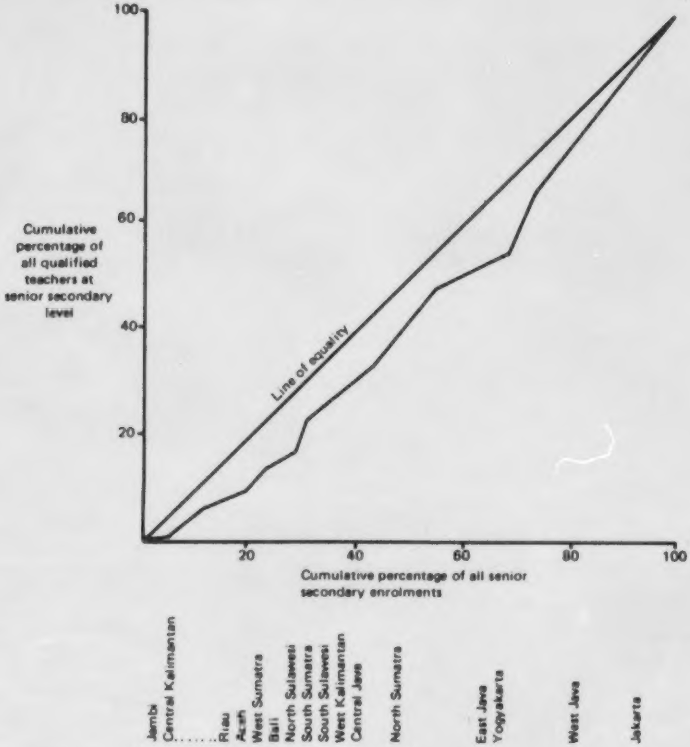


Diagram 8.1: *Lorenz curve for the distribution of qualified teachers at the senior secondary level across provinces – Indonesia, 1976*

The interchangeability of the indicators

Having calculated values for eight of the process indicators for each of the Indonesian provinces, the extent of the differences in the comment they make about the Indonesian provincial education systems must be assessed. If, for example, the indicators are at least statistically very similar, then sufficient comment can be made about the similarities and differences among the systems by the selection of just one of the eight indicators. The educational and conceptual uniqueness implied by each of the indicators as defined would then be lost. If this were the case, it would lower the credibility which could be placed on the applicability to provincial education systems of these indicators.

To test the extent of the similarity of the comment made by the indicators, a correlation matrix is calculated. This approach was recommended in Section 3.5 in the context of determining the interchangeability of indicators. The correlation matrix is reported in Table 8.9.

Indicator name	1	2	3	7	8	10	11	12
1. Student places	1.00	0.91	0.95	0.38	0.25	0.39	0.20	0.32
2. Schools	0.91	1.00	0.93	0.30	0.28	0.37	0.16	0.15
3. Classes	0.95	0.93	1.00	0.32	0.27	0.26	0.10	0.25
7. Female involvement	0.38	0.30	0.32	1.00	0.23	0.37	0.36	0.01
8. Orientation second-level	0.25	0.28	0.27	0.23	1.00	−0.08	−0.06	−0.12
10. Years of education	0.39	0.37	0.26	0.37	−0.08	1.00	0.75	0.06
11. Expected grade	0.20	0.16	0.10	0.36	−0.06	0.75	1.00	−0.01
12. Qualified teacher	0.32	0.15	0.25	0.01	−0.12	0.06	−0.01	1.00

Note: A value for the correlation coefficient, calculated from a sample of size 26, would have to exceed 0.45 to be significantly greater than zero at the 1 per cent level.

Table 8.9: *Correlation matrix for process indicators – Indonesian provinces, 1974*

The correlations among the three indicators measuring the availability of student places, schools and classes all exceed 0.90. These indicators are therefore measuring very similar characteristics of the provincial systems. This is an expected result and the similarity was anticipated when the indicators measuring school and class availability were defined in Section 5.2 to parallel the indicator for students. When using indicators to monitor the Indonesian system, therefore, it would not be desirable to use more than one of these three to make separate comment on development patterns unless particular policy implications are to be identified in two or all three of these particular areas. When this is the case, particular attention would have to be directed to those few provinces which were not ranked similarly on all three indicators.

The indicator measuring the expected number of years of education to be received correlates, with a value of 0.75, with the indicator measuring the expected grade level. This value is also very high and it reveals the considerable similarity of comment made by these two indicators. The similarity here is not as high as for the three indicators above, however, and so, despite the high similarity, the two indicators could, at least partially, be measuring slightly different characteristics. The selection of one or both of these inter-related indicators into a basic set of Indonesian provincial education system indicators would depend on the policy considerations of Indonesian planners and the objectives established for the provincial education systems.

All other indicators have very low inter-correlations. Although it is debatable as to whether a sample statistic is appropriate to use in this context, reference to a test of statistical significance does give a guide to interpretation. A correlation value of 0.45 would be required for statistical significance at the 1 per cent level. No correlation in the matrix exceeds this value except for those already mentioned. Hence it can be concluded that the remaining indicators are comparatively independent from each other and from the five indicators already discussed. A sub-set of five or six of the eight process indicators for which values are calculated for each province could therefore be used to monitor development patterns with some assurance at least that the characteristics measured by each indicator were not confounded by the characteristics related to the other indicators.

The demonstration of this relative independence is important not just for the Indonesian context but also for the development of indicators measuring characteristics of regional education systems. It demonstrates that indicators identified as being independent in a cross-national setting also have a degree of independence in this other context. Before this can be assumed with some degree of confidence, the independence needs to be demonstrated in other regional settings as well. The encouragement by the results in the Indonesian context is at least heartening and it supports the recommendations made tentatively in Chapter 6.

8.3 Résumé

A number of general comments need to be made about these demonstrations of the use of indicators to evaluate part of the first development plan for Saudi Arabia and to monitor the short-term development of education systems in the provinces of Indonesia. These comments pertain to the ways the indicators were used as well as to the implications each demonstration has for planners and researchers in the two countries.

In both situations, indicators were used to facilitate comparisons of a self-referenced nature. That is, the indicator values were independent of the values which might have been calculated for other education systems in the way they were estimated and in the way they can be interpreted. Although comparisons among these values were possible – and in the Indonesian context such comparisons were made – the main use of these self-referenced indicators is for measuring how much development occurs over a particular period in a single system. In the context of evaluation – such as was undertaken for Saudi Arabia – self-referenced procedures are very important. Planners do not always wish to plan patterns and levels of development for their systems which are related to those which prevail in other systems. Instead they wish to plan for their own particular needs in the context of the special characteristics of their own systems.

The difficulty of attempting only self-referenced comparisons is that, especially when development is being monitored, evolution might be so slow that a given system might simply not expand rapidly enough to allow the society which provides that system to improve at the rate it wishes. In such circumstances, norm-referenced comparisons are needed because they allow reference to a relative standard for a rate of growth. In the Indonesian context, the standard used for the norm-referenced comparisons based on index numbers was provided by the average development growth rates calculated over all provinces in the country. In this way, provinces which were developing more slowly than others could be identified.

What the demonstration has shown, therefore, is that both norm-referenced and self-referenced comparisons are important. They each have their own purposes and strengths and hence indicators incorporating these concepts should be utilized whenever appropriate. Extending this conclusion would encompass indicators which allow criterion-referenced comparisons as well. Unfortunately, as is noted in Section 3.4, few indicators of this type have yet been developed for education systems.

A second observation stemming from the demonstration of the usefulness of indicators is that, although the indicators have used a variety of mathematical bases, these differences do not confuse the evaluation or monitoring processes. Certainly the basis for each indicator must be specified explicitly. Given that this happens, however, the use of an index number to measure one characteristic and a Gini coefficient or a C index of concentration to measure another raises no problem in interpretation. Understandability and ease of interpretation are both preserved. This is an important observation because it is not possible to measure all characteristics of an education system with indicators using a single mathematical basis. That differenct bases can be used removes what potentially might have been a critical obstacle in the development of education system indicators.

A third observation is that, in the Saudi context at least, there was not a really close congruence between the statements of objectives and the definition of the indicators. As was pointed out in Chapter 1, this situation would prevail in most if not all education systems, for the statement of objectives is notoriously bad in all countries. Generally the objectives tended to be a little vague in specifying what the planned outcome was to be or they were too narrow and did not place a proposed expansion into the context of total development. Perhaps if indicators become more widely adopted by educational planners and administrators, such a situation as this will not occur very often. Planned development could then be specified at least partly in terms of which indicators are to be changed and by what amounts. Evaluation of progress could then be readily reported – even on an annual basis as part of the usual reporting procedures.

A final observation is that data unavailability prevented the estimation of values for a number of indicators posited in Chapters 5 and 6 as being useful for evaluating and monitoring educational progress. This was particularly true in the Indonesian demonstration. It must be acknowledged that planners and administrators have a very difficult task in determining what data they should collect from schools, teachers and pupils. For each system, different demands must be accommodated. This application therefore points out, in individual situations, particular data that are lacking for education system features which are considered as being important in those specific contexts. Similar applications in different countries would reveal the extent to which these and/or other data were lacking elsewhere. Through identifying system features by means of indicators, therefore, steps can be taken to have the appropriate data collected either through annual returns or through special surveys, etc. The adoption of a set of indicators thus helps to guide a rationalization of data collection procedures.

Chapter 9

The Use of Indicators in Researching Education Systems

The contribution indicators can potentially make to researching the nature and pattern of education system development is noted in Chapter 1. They can, for example, be used to measure system development itself or to represent educational provision and development level when the inter-relationship among a number of societal sub-systems is measured. To demonstrate one way in which indicators might be used to develop theoretically oriented research results, the context of the second of these research problems is selected. In particular, the question to be analysed is:

> What are the relationships among selected education system characteristics and other selected characteristics of society?

This problem has already been investigated by many different researchers, some of whom are listed in Table 1.1. The importance of the problem has thus already been convincingly established by many writers. Unfortunately no clear or consistent explanation of the inter-relationships has been forthcoming from these research efforts. One possible reason for this, and the characteristic common to all past research studies, is that individual variables have been used to represent the level and degree of education system development and the level and degree of development of the economic, political and social systems. The consequence of such an approach is that very particular characteristics from each type of system are related – and generally only in pairwise comparisons – and thus a very myopic view of the actual relationships among the characteristics results.

The potential advantage of using indicators to represent education system characteristics has been stressed many times in the preceding chapters. To be consistent, indicators should also be used to represent the characteristics of the other societal systems. By inter-relating sets of indicators derived from these various systems, the role an education system plays in a society can be placed in a much broader and clearer perspective.

9.1 The selection of indicators to measure societal characteristics

Difficulties immediately arise in seeking to identify the characteristics of society which should be selected for such an analysis as is proposed here and the ways in which these characteristics should be measured. Some research has been conducted in examining the multivariate inter-relationships among large numbers of variables measuring particular characteristics of society. Studies by Cattell (1949 and 1950), Berry (1960), Banks and Gregg (1965), Russett (1968) and Rummel (1972) are the principal examples of these. They all show that there are certain identifiable dimensions either within particular societal systems or across a number of these systems. Indeed, as both Russett and Rummel point out, their results show considerable similarity to the results obtained by the earlier studies. Certainly slight variations do exist but the factor analytically identified dimensions do, in general, point to particular yet similar societal characteristics as being those which most parsimoniously describe and differentiate among nations. These dimensions can therefore be adopted as empirically defined indicators and can be inter-related with sets of education system indicators so as to appraise and/or test hypotheses concerning the role of education in society.

Unfortunately few of the studies which have been done report values for their factors. Many indeed factor analyse incomplete data matrices and this complicates the calculation of such values for each nation. The study by Russett does however report scores for all nations on the factors identified. This study is therefore chosen to define the societal characteristics analysed here.

Russett analysed 54 variables which describe most of the characteristics commonly encountered in other research studies. By factor analysing this broad cross-section of variables, he identified four dimensions which he labelled:

- ☐ economic development
- ☐ communist influence
- ☐ catholic culture
- ☐ intensive agriculture.

He concluded that these dimensions 'correspond quite closely to the factors that have emerged from a variety of other studies' (1968, 327). Such stability in results lends support to the adoption of Russett's results here.

Certainly it is possible to disagree with this selection on the grounds that other important aspects are not included. Be that as it may, but a number of important characteristics are identified here which have been hypothesized to relate to characteristics of education systems and/or to the level of system development. For the purpose of this demonstration, therefore, the Russett dimensions are accepted as being a set which adequately represent a variety of societal characteristics and it is acknowledged that other, at least equally important characteristics must be omitted.

Russett's data all come from his *Handbook* and they pertain to years 'in the late 1950s' (1968, 322). By relating his dimensions to the empirically defined education system dimensions for 1960 – as reported in Appendix B – an appraisal of the immediate effect of the characteristics embodied in the societal dimensions is possible. Longer term effects can also be appraised by relating the Russett dimensions to the empirically defined education system dimensions for 1969 (as reported in Johnstone [1977]).

9.2 The method of analysis

The statistical technique of canonical correlation analysis is used to assess the inter-relationships between the two sets of dimensions. The results of this form of analysis indicate both the extent of the relationship between the two sets as well as those dimensions from each set contributing to the formation of the identified relationship.[1] Only those canonical correlations significant at the 5 per cent level and their corresponding dimensions with vector loadings greater than 0.4 are reported. To maximize the base for the analysis, all countries common to the two sets of dimensions being correlated are included. This means that a slightly different set of nations form the canonical correlations between Russett's dimensions and the 1960 input dimensions than form the correlations between the Russett dimensions and the 1960 process dimensions. This does not affect the interpretations made of the results. It does however increase the sample size quite significantly for each analysis over what it would have been had only those countries common to all dimension sets been used.[2]

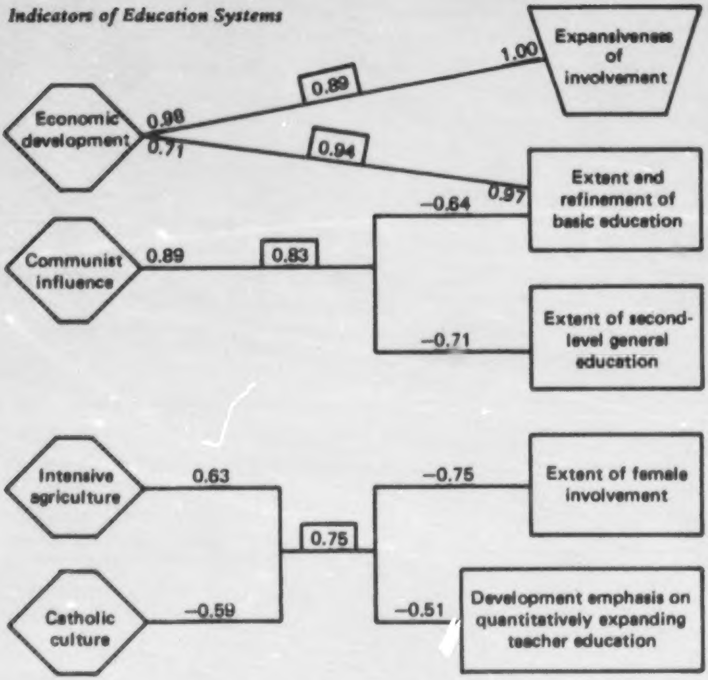
9.3 The short-term relationships

The short-term results are summarized in Diagram 9.1 while the specific details of each correlation are reported in Table 9.1.

Both the single significant correlation between the Russett-1960 input sets and the highest correlation between the Russett-1960 process sets identify countries having developed economies in the late 1950s as also having, in 1960, the most expansive involvement in the provision of educational facilities and the most extensive and refined basic

---

1. For a description of the technique of canonical correlation, see Cooley and Lohnes (1971), Kerlinger and Pedhazur (1973) and Darlington *et al* (1973).
2. The value for Denmark reported by Russett on the 'communist influence' factor has been changed because it appears to have been misprinted. The value used is 0.84 and not 0.084 as reported.


Economic development
Communist influence
Intensive agriculture
Catholic culture
Expansiveness of involvement
Extent and refinement of basic education
Extent of second-level general education
Extent of female involvement
Development emphasis on quantitatively expanding teacher education
0.98
0.71
0.89
0.94
1.00
0.97
0.89
0.83
–0.64
–0.71
0.63
–0.59
0.75
–0.75
–0.51

Legend: Values enclosed thus ☐ are canonical correlations between indicators joined as indicated. Values next to indicators are loadings on corresponding canonical vectors.

Diagram 9.1: *Inter-relationships between the Russett and the 1960 educational indicators*

Indicator combination	Canonical correlation number	Correlation value	χ^2	% of variance
(a) Short-term relationships				
Russett – 1960 input	1	0.89	88.79	79.8
Russett – 1960 process	1	0.94	82.10	88.8
	2	0.83	42.89	7.6
	3	0.75	31.04	2.0
(b) Long-term relationships				
Russett – 1969 input	1	0.86	77.94	74.5
Russett – 1969 process	1	0.95	113.20	89.4
	2	0.79	48.84	6.6
	3	0.62	24.08	1.5

Table 9.1: *Summary results of canonical analyses between the Russett indicators and the 1960 and 1969 sets of education system indicators*

educational provision. This is in broad agreement with the results of previous studies (eg by McClelland [1966], Edding [1962] and Harbison and Myers [1964]) but on a much broader scale. It appears then that countries in the best economic position to provide education, do so.

Diagram 9.2 plots the values for each country on each of the canonical variables associated with the main Russett-1960 process correlation. It shows that countries such as Argentina, Honduras and Panama, although not having very extensive basic educational provision, nevertheless do have more extensive provision than their economic development level indicates. On the other hand, countries such as Italy, USA and Venezuela have less refined provision then might be expected to be associated with countries at their respective levels of economic development.

The second canonical correlation between the Russett factors and the 1960 process dimensions has a value of 0.83. The vectors associated with it show that the communist countries have not extended or refined their basic education and they have few enrolments in the second-level

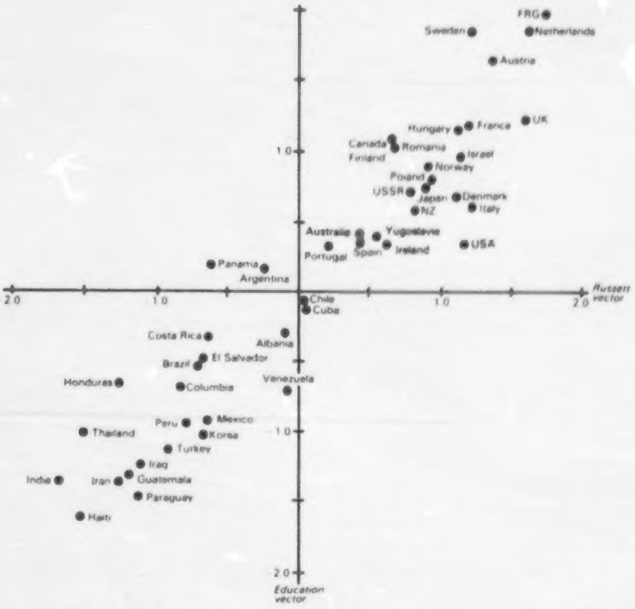


Diagram 9.2: *Canonical vectors relating the Russett 'Economic development' indicator to the 1960 'Extent and refinement of basic education' indicator*

general education stream. This appears to be an idiosyncratic characteristic of a group of countries which includes some from Eastern Europe together with Argentina and Israel. The plots of the values of the canonical variables for this correlation in Diagram 9.3 indicate that although Argentina and Israel have this similarity in their educational provision, their totally different political systems clearly separate them from the communist countries. The diagram also shows that non-communist countries such as Japan, UK, South Korea and Thailand strongly emphasize enrolments in the second-level general education stream while the first two also have more refined and extensive basic education.

This inter-relationship shows that the communist countries do have a distinctive pattern in the process aspects of their educational provision. If technical education is viewed as producing workers who can be used for production purposes rather than individuals who principally pursue their own interests, then one of Noah and Eckstein's hypotheses is supported. They postulate: 'School systems in communist

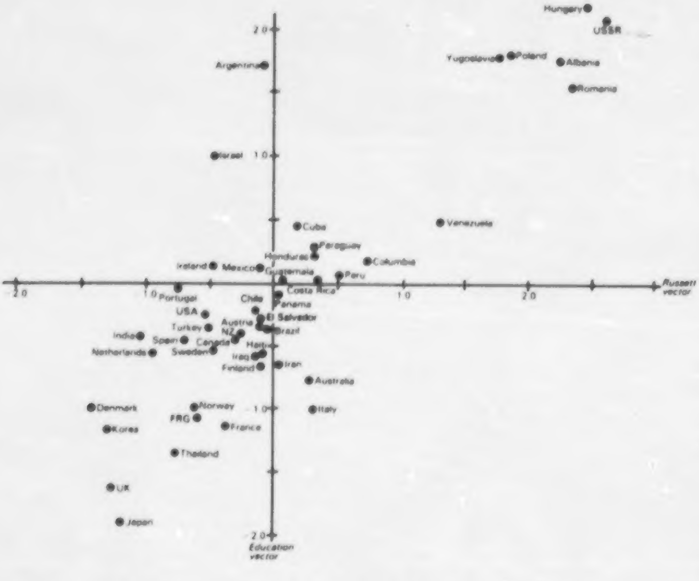


Diagram 9.3: *Canonical vectors relating the Russett 'Communist influence' indicator to the 1960 indicators measuring the 'Extent and refinement of basic education' and the 'Extent of second-level general education'*

countries are designed to serve the interests of the state more than are schools in non-communist countries' (1969, 206). The communist countries do provide extensive pre-school and special education and if these facilities also are viewed with ulterior motives in mind, those aspects of the refinement of basic education indicator can be viewed as serving 'the interests of the state'. However, in the absence of proper definition of what constitutes state interests to the exclusion of the needs of individuals (as is included in another of their hypotheses – number 10), the present study merely concludes that communist countries do differ from almost all non-communist countries on two of the process indicators.

The third and final significant canonical correlation between the Russett and the 1960 process dimension sets explains a very small proportion of the total variance yet it is still statistically significant. It describes countries having an intensive pattern of agriculture[1] as also having particularly low female involvement in education and a low emphasis on developing teacher education facilities. The 'catholic culture' dimension, which Russett describes as measuring 'a fairly broad cultural phenomenon rather than a purely religious one' (1968, 326) is also low for such countries.

The country values on the associated canonical vectors are plotted in Diagram 9.4. Many of the Latin American countries are located in the lower left quadrant because the antithesis of the above country description is appropriate for many of these countries. The top right quadrant contains both economically developed and developing nations. All countries located here have extensive agricultural development (as defined by Russett's dimension), catholicism within their societies is generally fairly low, they have fewer females involved in education than indicated by world averages and they are slowly increasing their teacher education facilities.

This result modifies the conclusions of previous studies investigating the relationship between economic development and education. Such studies use gross variables to represent these two societal sub-systems. Then, upon finding a high relationship between the two variables, the researchers conclude that educational and economic development are inter-related in some way. This finding, however, merely indicates that highly advanced economies support extensive education systems overall. The modification made by the results reported here is that, if a country

---

[1] The 'intensive agriculture' dimension is not well labelled as Russett readily admits: 'variables loading highly on it include overall population density and density as related to agricultural land and, negatively, area. One might call the dimension simply "density", except for the fact that inequality of land distribution is rather negatively correlated with it, emphasizing that in sparsely settled countries there are not only small family farms but a number of great estates as well' (1968, 326-7). Despite the poor descriptive label, this study accepts the Russett terminology for purposes of description.

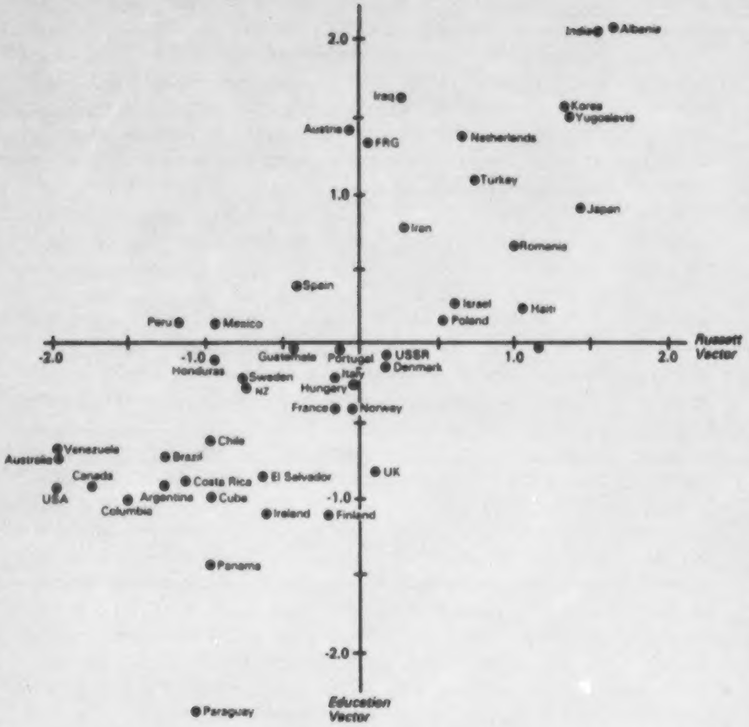


Diagram 9.4: *Canonical vectors relating the Russett 'Intensive agriculture' and 'Catholic culture' indicators to the 'Extent of female involvement' and 'Development emphasis on quantitatively expanding teacher education' indicators for 1960*

is not catholic in orientation, is small, densely populated, and has only a small proportion of its work-force in the agricultural sector, its provision for females is low – even though its overall educational development level may be high. Hence, developed economies which are classified by Russett as having 'intensive agriculture' – such as Germany, Netherlands and Japan – have extensive education systems particularly for basic education, but they do not have equivalent educational involvement for both females and males. Low level economies which have a catholic orientation, and which are large, more sparsely settled and less dependent on agriculture – such as Columbia, Panama and Paraguay – although not having extensive basic educational provision, have at least attempted to make adequate provision for females. Low level economies, which are also dependent on agriculture but are not catholic – such as India, Korea and Turkey – do not have extensive basic educational provision for either males or females.

9.4 The long-term relationships

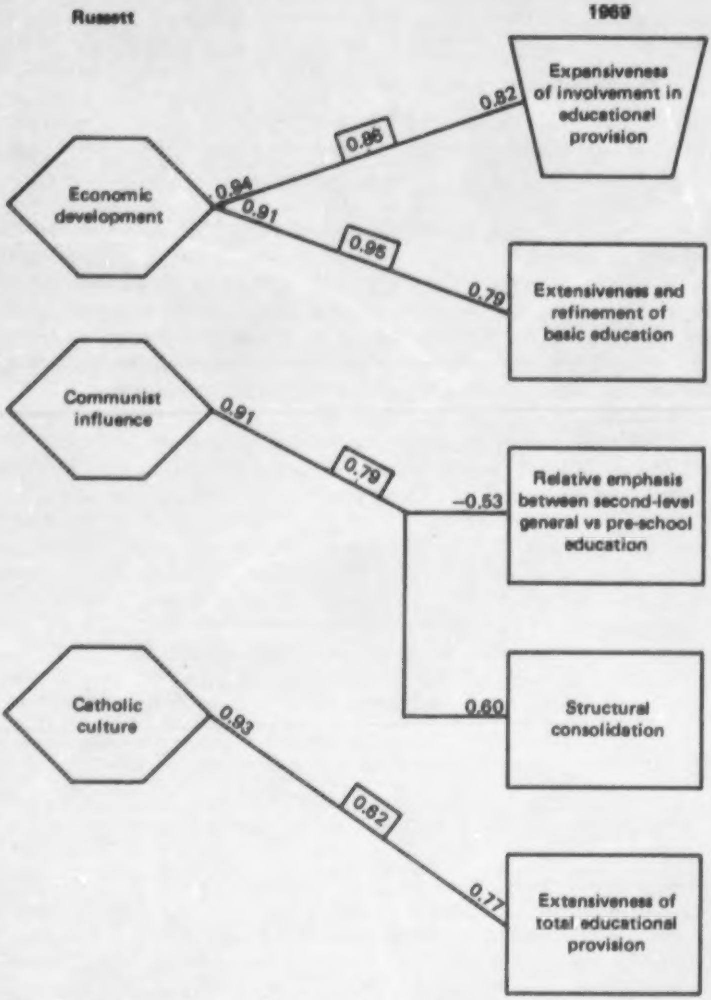
The results of examining the long-term effects of the characteristics measured by Russett on education system provision are summarized in Table 9.1 and Diagram 9.5. They are fairly similar to those obtained for the short-term analysis. Those countries which, in the late 1950s were the most economically advanced have the most expansive involvement in educational provision in 1969 and they have extended and refined their basic educational provision. The relationship identifying countries with high communist influence as also tending to have extensive pre-school provision also remains. The 1960 relationship also incorporated the 'basic education' indicator. The present relationship measures this latter characteristic with the 'structural consolidation' indicator. Such a substitution is understandable because of the general worldwide movement to improve and consolidate all aspects of basic education.

The third significant canonical correlation is between the catholic culture indicator and the extensiveness of total educational provision indicator. In the Russett-1960 analysis, these two are related through the corresponding 'female involvement' indicator although there the relationship is confused by other factors as well. The vectors associated with this 1969 correlation – graphed in Diagram 9.6 – again clearly distinguish the Latin American countries having both high female involvement in education as well as high catholic orientation.

Another of Noah and Eckstein's hypotheses is tested by both the short-term and the long-term analyses reported here. They postulate:

> among christian countries, the more catholic the country, the lower its educational level. (1969, 207)

What is meant by the term 'educational level' is unclear. Diagram 9.5 shows clearly that there is a positive relationship between catholicity of culture in general (and catholic religion in particular) and high female involvement in education. No negative relationship between catholic culture and either extensiveness of educational provision or extent and refinement of basic education is determinable. Although the definition of 'catholicity' is not exactly what Noah and Eckstein intend, there is little support for their postulated relationship but a great deal of support for its antithesis phrased with a more specific orientation to female involvement.


Russett
1969
Economic development
Communist influence
Catholic culture
Expensiveness of involvement in educational provision
Extensiveness and refinement of basic education
Relative emphasis between second-level general vs pre-school education
Structural consolidation
Extensiveness of total educational provision
0.94
0.86
0.82
0.91
0.95
0.79
0.91
0.79
–0.53
0.60
0.93
0.82
0.77

Legend: Values enclosed thus ☐ are canonical correlations between indicators joined as indicated. Values next to indicators are loadings on corresponding canonical vectors.

Diagram 9.5: *Inter-relationships between the Russett and the 1969 educational indicators*

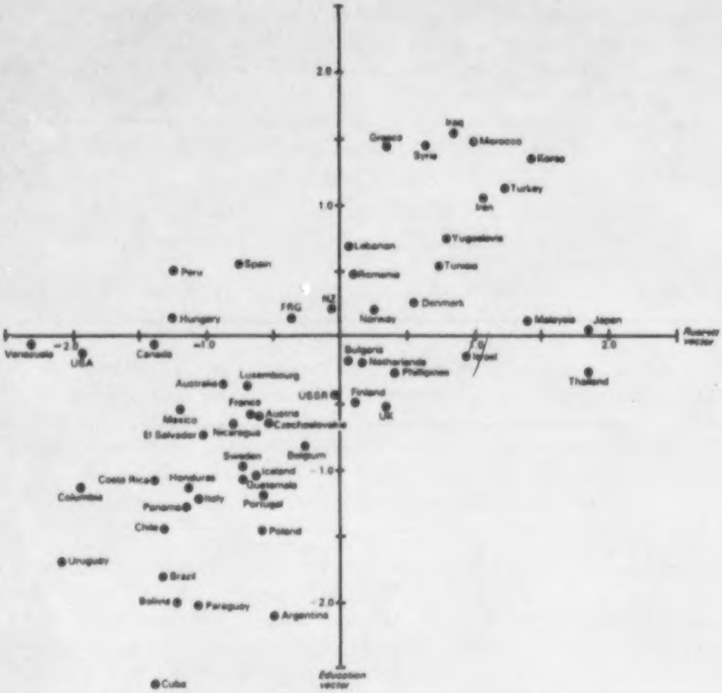


Diagram 9.6: *Canonical vectors relating the Russett 'Catholic culture' indicator to the 1969 'Extensiveness of educational provision' indicator*

9.5 Résumé

This use of the education system indicator sets to analyse relationships among selected societal characteristics and aspects of national education system provision is in no way intended to be a complete summary of results which may be derived. Certainly the results have interest and implications for educational planners and researchers. The main point in presenting them however is to demonstrate one way in which indicators might be used in researching education systems development. The analyses thus present an example of the kind of approach which may be followed. They also demonstrate the increase in the power of analysis when use is made of a number of indicators rather than of single variables as has been the practice so often in the past.

The inter-relationships between Russett's dimensions on data relating to 'conditions in the late 1950s and the 1960 and the 1969 education dimensions respectively correspond very closely. The actual differences are explained either in terms of observed shifts or factor substitution although the canonical correlations tend to be less in value over the longer period. The results show clearly that certain political, economic and religious dimensions of some societies are related directly to characteristics of the education systems within those societies. In particular, the following hypotheses can be said to have been supported:

1. Levels of economic development and the extensiveness of educational provision in a country are intimately and positively inter-related.
2. Countries having intensive agriculture do not provide the same opportunities for females in their education systems as they do for males.
3. Communist countries have a distinctive pattern of educational provision. However, possession of that pattern by a country does not imply a country has extensive communist influence.
4. Countries with a high catholic culture have relatively high female involvement in education almost independent of either the level of their economies or the extensiveness of their education systems.

Continued research to increase this list would contribute to theory building in the way described previously. Such a theory would have significant importance for the understanding of the total development process as well as for predicting the effects of policy decisions advocated in the process of educational planning. To accomplish this, however, the most important and influential indicators of the various characteristics of societies need to be defined and/or identified and then developed to a proper level of sophistication.

Chapter 10

The Use of Indicators to Classify Education Systems

Another purpose identified in Chapter 1 as motivating the use of indicators is the need to classify education systems. Some of the reasons for such classification have already been discussed. In the planning context, typologies can reveal those education systems which are alike and those which are different across a broad range of education system indicators. Such typologies can be developed either in a cross-national setting or among regional education systems in a particular country. The extent of the similarities and differences among education systems can also be assessed. Development statuses and capacities of various countries or regions within a country can thus be identified analytically as too can what targets might be feasible and appropriate for adoption by particular developing education systems. The results of implementing policies to change system provision over one or two planning periods might also be assessed in either a norm-referenced or a criterion-referenced way.

In the research context, many writers have advocated the formation of country typologies as a possible means of aiding theory development.[1] Stern *et al* (1973) claim that there is evidence of such aid in fields such as politics. They claim that the typologies established by Almond (1956) and Almond and Verba (1963) formed a basis for later research aimed at developing theories of political culture. Indeed Stern and his colleagues also attempt to build upon the earlier results.

Commenting generally on research approaches, Anderson asserts that the use of typologies and the resulting 'comparative method is one of the numerous imperfect substitutes for experimentation' (1961, 5). Typologies are the closest that a planner or a comparative education researcher can get to experimental groups. They serve as a substitute in a situation where it is impossible – practically as well as both morally and politically – to manipulate education systems in order to determine the effect of a particular policy decision (ie experimental treatment).

---

[1] Such writers include Berry (1960, 78); Anderson (1961, 14); Bereday (1964, 110); Beeby (1966, 50); Berstecher and Dieckman (1969); and Welsh (1971, 274).

By comparing growth and development in countries from different groups within a typology, a first step towards generalizing research results can be taken. Further results can then be interwoven in an identifiable way into the earlier studies.

In this chapter, two distinct forms of typologies are formed. The first identifies groups of national education systems which are similar to one another. The second identifies homogeneous groups of provincial systems within a single country. These situations are equally appropriate to both the planning and the research uses of education system indicators.

10.1 Basic approaches to typology formation

Previous attempts to form typologies of education systems can be divided into three classes according to the type of approach used. The first approach involves the mere acceptance of a hypothesized typology but not its demonstration or validation. Such an approach is used, for example, in the Karachi Plan (see Unesco [1966]). There countries in the Asian region are classified into one of three groups 'with reference to the time in which they are likely to achieve at least seven years of universal compulsory education, namely: (A) those which are likely to achieve it after 1980; (B) those which are likely to achieve it around 1980; (C) those which are likely to achieve it before 1980' (Unesco, 1966, 39). This approach therefore makes assumptions about possible educational development patterns but it does not explicitly depend on education system indicators. Consequently it results in a hypothesized typology without a sound empirical foundation. Although such a typology can be useful for some purposes, it is not discussed further here.

A second approach uses an *a posteriori* and descriptive method of defining the typology. Studies classified under this approach have the characteristic that the researchers postulate some basic typology of educational provision. This plan then forms the basis for discussion about either education systems *per se* or the relationship between educational provision and some other characteristic such as the process of modernization.

Typical of such studies are those by Beeby (1966) and Hopper (1968). Beeby properly admits his is a non-empirical study and that his thesis is based largely on his own (very extensive) observations of world education systems. He posits four stages in the development of an education system – each stage related to the level of education and training of teachers. Although in his discussion, Beeby alludes to some broad characteristics of education systems and describes the change in type of provision in some countries, this is by way of example rather than as either a demonstration of the use of indicators or a test of his typology. He never attempts a classification of countries, for example, according to his hypothesized framework.

Hopper takes a quite different approach. He works from Turner's (1960) article where a typology is hypothesized according to the type of upward social mobility conferred on individuals by the education system in a given country. Hopper considers that the two types of mobility defined by Turner – contest mobility and sponsored mobility – do not distinguish adequately between the systems of many of the world's leading countries. He therefore expands the typology to include consideration of the structure of the selection processes incorporated into education systems. He then uses his typology to compare, contrast and discuss the different educational patterns manifested in eight countries.

There are two main weaknesses of such research studies as these. One weakness is that few variables let alone indicators can be accounted for at any one time in the typology. As education systems have a multitude of components, a restriction on the number of variables must limit any investigation to considering one specific aspect of total provision – such as Beeby's restriction of his approach to teacher education. At times – and Beeby's situation would be one – this may be a perfectly acceptable limitation. On other occasions, however, it is equivalent to an inference that two people are alike because they both have black hair. Particularly if systems are to be compared for purposes of broad systems planning or for theory development, a wide range of educational features – and therefore indicators – must be used.

A second weakness of the approach is that it lacks statistical rigor. Some people may see this as a strength of the approach. Without using some sorts of statistical scales and statistical techniques, it is however difficult to describe and/or classify many systems simultaneously or use many indicators so that more than obvious similarities and differences among systems can be noted. If, rather than employ such techniques, a typology is limited to only a few systems, then its usefulness is also restricted.

In the category of analytical and statistical approaches to typology construction, the field is all but deserted. One study which does use an elementary statistical approach is that by Harbison and Myers (1964). On the basis of their composite index already described in Chapters 3 and 5, they divide 75 countries into four groups and term these:

☐ the underdeveloped countries
☐ the partially developed countries
☐ the semi-advanced countries
☐ the advanced countries.

There are between 16 and 21 countries in each group. The researchers then extensively review patterns of educational provision within each group.

The main reason why few research studies aim to form typologies of education systems is probably that the usefulness of typologies for further research has not been fully realized by empirically oriented researchers. As indicators of education systems have been developed and shown to proffer useful comment on education system development, they are now used to demonstrate ways in which a typology might be constructed. The potential usefulness of the typology can then be assessed.

10.2 Methods of empirically forming typologies from a set of indicators

Empirical methods which can be used to form typologies from a set of indicators are classifiable into four main groups:

- ☐ normal pattern similarity coefficients such as those reviewed by Sokal and Sneath (1963)
- ☐ the McQuitty type clustering methods depending mainly on grouping entities according to their inter-correlations (see, for example, McQuitty, 1966 and 1971)
- ☐ Q technique factor analysis (see Cattell, 1952 and Rummel, 1970)
- ☐ methods estimating distances between entities in m-dimensional space (where m is the number of indicators being used for the comparisons).

A combination between the nature of the data of most education systems indicators and the nature of the problem determines that the first three of these methods are not appropriate to use here. The method of grouping based on euclidean distances is therefore adopted. This approach is described in detail elsewhere (Johnstone, 1976), and so only a brief overview of the concept of the clustering approach is needed here.

The method begins by examining the profiles drawn across a series of defined indicators for a number of education systems. A typical profile of this type is drawn in Diagram 1.2. The similarity between any two profiles – that is between any two education systems – is then estimated by calculating the overall distance between the two profiles. The smaller this overall distance, the more alike two profiles are. Hence, if the overall distance between profile A and profile B is calculated to be greater than that between profile A and profile C, then it may be concluded that system A is less like system B than it is like system C.

To estimate the overall distance between two profiles drawn across a set of education system indicators, the actual distance between corresponding pairs of indicator values in the profiles for each of the two systems must be used. In mathematical terms, the estimation is computed from the formula:

$$D_{ij}^2 = \sum_{k=1}^{m} (I_{ki} - I_{kj})^2 \qquad \textit{(Equation 10.1)}$$

where D_{ij} is the distance between system i and system j
m is the number of indicators (I_{ab}) across which systems i and j are being compared.

The typology is formed by clustering together those systems whose profiles are most similar – that is where inter-profile distances (or D^2 values) are smallest.

Of the many algorithms which can be used to form the clusters of systems from the inter-profile distances calculated between all possible pairs of systems being considered, the one known as the 'diameter' or the 'centroid' method is the one recommended as being preferable. Its application is demonstrated in Appendix C. Basically this algorithm attempts to form groups such that each group has maximum internal homogeneity in space – that is, each group is contained within a sphere with the smallest possible diameter. Phillips concludes that the diameter method is most appropriate 'when a researcher wishes to break-up a dense cluster of entities' (1968, 30). As the profiles representing education system characteristics for national or regional systems do tend to be very similar in certain ways, the formation of groups having maximum internal homogeneity for any given inter-profile distance (or D^2 value) appears to be a realistic solution to typology formation.[1]

One feature which can complicate the clustering of system profiles is that three sources of variation contribute to the D^2 statistic as calculated by Equation 10.1. These sources are:

☐ the average level of the profile across all indicators
☐ the scatter of the observations about that average level
☐ the shape of the profile.

Slightly different clusters of systems can sometimes result by considering one rather than another of these properties. For example, the two profiles drawn in Diagram 1.2 would cluster together in a typology based on the average levels of profiles across all indicators but they would not cluster in a typology based only on profile shape. Hence a preferable first step is the formation of two separate education system typologies for any given set of indicators – each typology being based on a separate profile property. The final typology can then be determined by eliminating systems not clustering together in both sets. Only those systems which are most alike are then classified into the same final cluster.

Of the two typologies to be formed, one should be based on D^2

---

[1] **Full technical details of the diameter method together with an analysis of its effectiveness and efficiency are provided in Johnson (1967).**

values calculated by Equation 10.1. This typology is known as the unadjusted typology. It considers all three properties although by far the most dominant component in the values calculated is the average level of the profile. The other typology should be based on only the estimated similarities in profile shape. It is called the adjusted typology. The inter-profile similarity measure here adjusts the value of D^2 as calculated by Equation 10.1 by removing the components related to average level and scatter. The equation describing this adjustment is:

$$D^{*2}_{ij} = \frac{D^2_{ij} - m\Delta^2_e - (\Delta S_{ij})^2}{S_i S_j} \qquad \textit{(Equation 10.2)}$$

where D^{*2}_{ij} is the adjusted distance between system i and system j
D^2_{ij} is the distance between systems i and j as calculated by Equation 10.1
Δ^2_e is the difference in the average indicator value between the profile for system i and that for system j
S_k is the standard deviation of the profile for system k
ΔS_{ij} is the difference between the standard deviation for system i and that for system j.

Two examples are now presented to demonstrate how typologies of education systems can be formed from some of the indicators defined earlier. The first of these examples demonstrates typology formation in a cross-national setting. The profiles across four input indicators for each of 85 nations are compared to determine which education systems are most similar. The second example demonstrates typology formation in an intra-national setting. Here provinces within a single country (Indonesia) are compared across a set of process indicators to determine the extent of similarity of educational provision in different regions within that country.

10.3 Cross-national typology formation from input indicators

To demonstrate the formation of a cross-national education system typology, four indicators are selected from a set of input indicators summarized in Table 5.2. The ones chosen are:

☐ involvement of human resources in the provision of education
☐ development emphasis on changing human resource involvement
☐ financial involvement
☐ financial expansion.

Data for the component variables for each of these indicators are available from the Unesco *Yearbooks* for many countries. They are not however so freely available for so many countries for the remaining three indicators listed in Table 5.2. A sample of 85 developed and developing countries is chosen and the data used relate to the year

1969.[1]

For each country, the component variables of each indicator are standardized, summed and averaged according to Equation 3.4. The reason motivating this operation is that the indicator values used to form the clusters of education systems should have approximate mean and standard deviation values of 0.00 and 1.00 respectively. Comparisons between the values for the indicators are then facilitated and no one indicator is made more influential in determining the clusters than any other.

From the indicator values formed, the D^2 statistic is calculated[2] using Equation 10.1 for every pair of systems. Each national education system therefore has 84 estimates of the similarity between its profile across the four input indicators and the profile for every other system being analysed here. These values are then compared (in the way described in Appendix C) to form the unadjusted profile typology. The result is presented diagrammatically in Diagram 10.1. Continuing the clustering process beyond the point shown would have continued to enlarge the clusters but it would also have made the clusters less and less homogeneous. A decision on a particular stopping point can be made by noting practical considerations or the relative homogeneity within clusters compared to the heterogeneity among clusters or by merely inspecting the D^2 value and noticing when a large discontinuity occurs. These last two features are relatively interchangeable.

Of those systems that are shown as joining any cluster, the penultimate system to link is Ceylon. The value for the error (like the D^2 value) at which this occurs is 1.08. The four previous values at which links had occurred all exceed 1.00 and hence there is considerable similarity between the type of similarity being measured around this point. Then, however, Tunisia links at 1.25. The difference between 1.08 and 1.25 thus reveals a significant discontinuity and hence it signals that the clustering should stop here.[3]

Of the 85 national systems analysed, 78 form into six clusters. One of these clusters is particularly large and contains 49 national education systems. At the other end of the scale, there are two clusters which each contain only two systems. The systems which do not join any cluster are for the nations of:

Chad Cuba Ghana Greece Ivory Coast Madagascar Nepal

The membership of the typological clusters in Diagram 10.1 is confounded by the clusters being formed mainly on the basis of

---

[1] Criteria for country selection relate to data availability, reliability and validity. A fuller listing of the criteria used here is given in Johnstone (1977, 53-4).

[2] In this case, the value for m in Equation 10.1 is 4.

[3] If the clustering had been continued, the next stage would allow Nepal to join the Afghanistan cluster. The error value would be 1.79 – an even greater discontinuity than that depicted for Tunisia in Diagram 10.1.

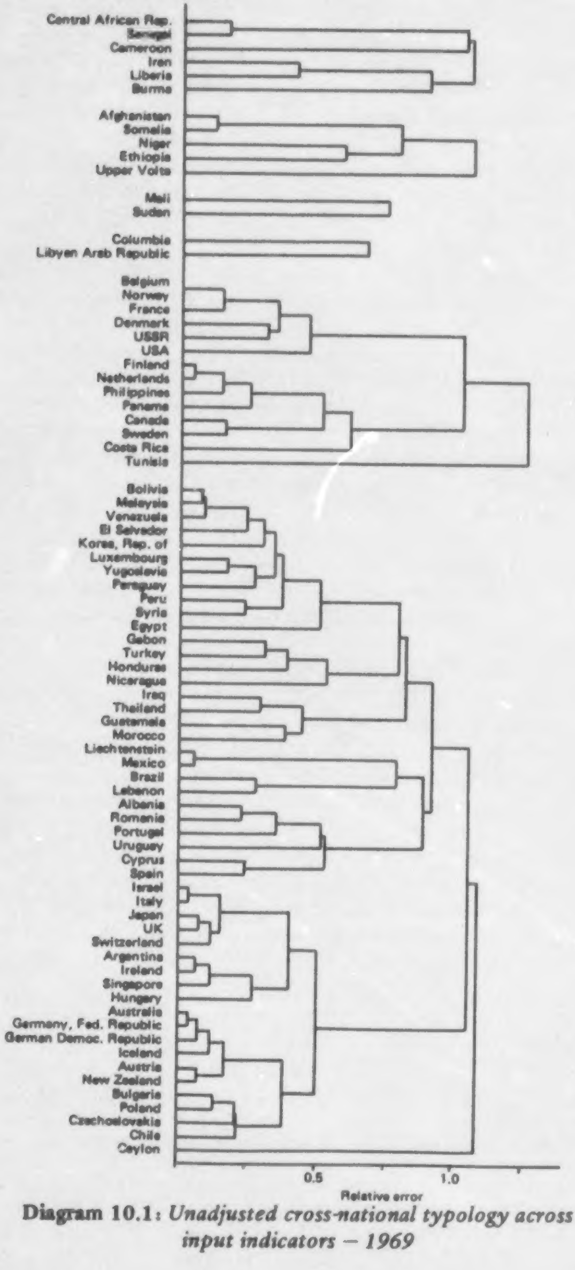


Diagram 10.1: *Unadjusted cross-national typology across input indicators – 1969*

estimating which systems have similar average values across all four indicators. As noted in the previous section, comparing the shapes of the profiles in an adjusted typology is important so that differences between sub-groups of the unadjusted clusters can be identified. Hence the procedure as described to this point must again be repeated but using the adjusted D^2 statistic as described by Equation 10.2 as the basis for comparing the national education systems. This second application of the procedure produces 19 clusters covering 74 countries. Eleven education systems are left not clustered – the implication of this being that the profiles of these systems across the four input indicators are exceptionally idiosyncratic with respect to the 85 systems analysed here. The national systems not clustering in the adjusted typology belong to:

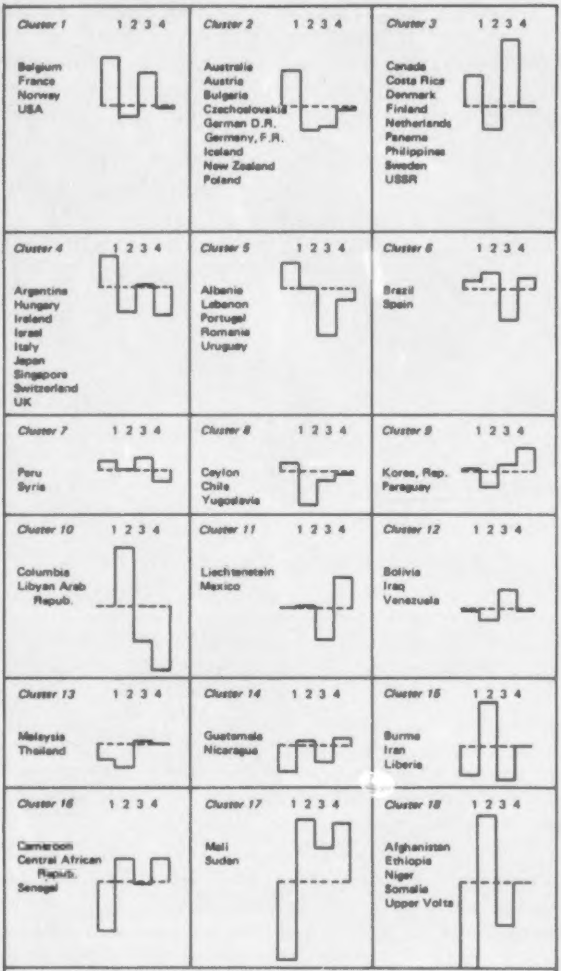
Cuba	Cyprus	El Salvador	Gabon
Greece	Honduras	Luxembourg	Madagascar
Morocco	Tunisia	Turkey	

Identifying the national systems included together in the same groups in both the unadjusted and the adjusted typologies reveals those systems which are similar in their overall level as well as those with similarly shaped profiles across the four input indicators. For example, although six nations combine to form the first of the unadjusted clusters shown in Diagram 10.1, three of them – Central African Republic, Cameroon and Senegal – form into one cluster in the adjusted typology while the remaining three nations – Burma, Iran and Liberia – form into another. Hence two distinct groups are identified. Dividing clusters into smaller units in this way as well as eliminating national systems not clustering together consistently (eg Tunisia) produces 18 distinct patterns of inputs into national education systems. These groups of systems, together with their average profile[1] across the four indicators are presented in Diagram 10.2. The clusters are drawn in order of a decreasing average value for indicator 1.

The profiles show considerable differences among clusters. For example, systems in the first cluster show high levels of involvement of human resources (indicator 1), below average rates of increasing this involvement (indicator 2), high financial involvement (indicator 3) and about average rates of increase in the relative flow of finance into their education systems (indicator 4). Systems in the second cluster manifest some similarities to these characteristics although their overall levels are not quite as high. The main distinguishing feature between cluster 1 and cluster 2, however, is that the level of financial involvement in providing

---

[1] Seventy of the 72 standard deviations for the profiles in Diagram 10.2 are less than 0.30 and none exceeds 0.45. These values compare with a value of 1.0 which is the approximate value for each of the four indicators across all 85 systems. These small values show the effectiveness of the clusters formed in the typology.



Notes:

1. Units are as indicated in cluster 1 and are standard deviations.
2. The numbers above each profile refer to the number of the indicator as follows:
   1. Involvement of human resources in the provision of education;
   2. Development emphasis on changing human resource involvement;
   3. Financial involvement; 4. Financial expansion.
3. Countries analysed but not joining any cluster are:

Chad	Cuba	Cyprus	Egypt
El Salvador	Gabon	Ghana	Greece
Honduras	Ivory Coast	Luxembourg	Madagascar
Morocco	Nepal	Tunisia	Turkey

Diagram 10.2: *Final cross-national typology based on input dimensions, 1969*

education in the systems in cluster 2 is comparatively quite low. By examining the other profiles, the reason why the Central African Republic group split from the Burma group (as noted above) can be clearly seen.

Planners and researchers with different backgrounds and who come from different environments would interpret the implications of the clusters forming the typology in Diagram 10.2 differently. Partly because of this as well as a respect for the validity of the multitude of interpretations which are possible, no interpretation of particular features of the individual clusters is made here. Individual planners and researchers are thus encouraged to discuss and proffer their own interpretations on the implications they see as being derivable. Instead, three comments are made here about the use of the typology.

Firstly, the typology should be re-calculated for a later year – say 1979 or 1980 – across the same four input indicators. Differences among development patterns could thus be identified. National systems following different development paths could then be examined to determine what policies might have led to these different patterns. The effect of those policies could therefore be evaluated. Conversely, systems classified into the same clusters in each analysis could be examined to determine if they in fact did have similar education policies. If they did not, reasons for the similarity between their development paths with respect to their inputs could be hypothesized and tested.

For example, although the systems in Ceylon, Chile and Yugoslavia are slightly more extensive in their involvement of human resources than those in the Republic of Korea or Paraguay, both Korea and Paraguay are increasing their financial investments in education at very rapid rates. The long term consequences of this difference should by now be apparent. Planners in these five countries should therefore be assessing what differences have resulted in their respective input levels and what changes have occurred in the patterns of their process indicators. Similarly, the development levels in Columbia and the Libyan Arab Republic can be seen to be comparable to those existing in Leichtenstein and Mexico. The essential differences between the profiles depicted for each cluster are in the two development emphasis indicators. The first two countries appear to have been expanding their involvement of human resources but not providing additional funds at a very rapid rate. The other two countries were not expanding their human resources but they were rapidly increasing the amount of their financial involvement. A comparison of this type based on 1980 data might proffer comment on the relative importance of expanding systems either with additional human resources or with a significantly greater money supply.

Secondly, the typology presents education system researchers with an indication of which countries are similar to each other and which are different at least with respect to the input characteristics measured here. Hence a basis is formed for matching national education systems for a quasi-experimental investigation. This corresponds to matching individuals in the more traditional and conventional psychological-educational experiments. Such investigations could analyse a variety of hypotheses especially within the contexts of education systems evaluation and policy research.

Thirdly, planners in various countries can ascertain the types of changes which would be required if they adopted other countries as targets for their development plans. This possible use of indicators in planning education systems development was discussed in Chapter 1. For example, Spain might wish to adopt as a target for a development plan, the pattern of inputs which exist in Italy. This would appear to be a reasonable and realistic procedure although there are many cultural and political differences between the two nations.[1] The two development emphasis indicators for Spain are higher than those for Italy – thus indicating that Spain is in a good position to expand its education system inputs in the near future. The two indicators measuring the present level of involvement reveal, however, that the rate of increase for the financial involvement indicator must be much higher than the rate of increase for the human resource indicator. Similar comparisons between nations for the purpose of establishing targets reveal that Burma or Iran might consider choosing Iraq and that Guatemala or Nicaragua might consider levels in Paraguay as possible targets.

This presentation of a typology of national education systems has not aimed to discuss the implications of the typology nor the advantages of its construction. It aims only to show what such a typology is like and what comments might be made from it. Planners and researchers should experiment further with such typologies in their own settings.

One limitation of the typology is that it is based on only four of the seven input indicators. Incorporating the other indicators could alter the membership of some of the clusters formed here. In addition, the typology based on the process indicators needs to be established for the same set of national education systems. It would not necessarily be similar to the input typology because of the different nature of the characteristics measured by the process indicators. These additional tasks fall outside of the purpose of this book and are recommended to fellow planners and researchers.

---

[1] There is, of course, no implication in the adoption of targets that nations or their education systems will be transformed so as to resemble the target situation in all respects. Referring to the above targeting situation, Spain, for example, would not be transformed into a 'mini-Italy'. It is only the levels of the indicators which are adopted as targets – not the national characteristics embodied in a nation.

10.4 Intra-national typology formation from process indicators

As has already been pointed out in various contexts, it is sometimes useful for planners to group or cluster education systems considered at the state or the provincial level. Such an exercise reveals the extent of the homogeneity or heterogeneity which exists among provinces within a nation. This in turn reflects how applicable national policies might be across all provinces or whether, instead, specific policies are needed in different geographical areas of a country.

To demonstrate how such a regional typology might be formed and what it might be like, the 26 provinces within Indonesia are analysed on the basis of the 1974 process data summarized[1] in Section 8.2. Process indicators 2 and 3 in Table 8.7 – measuring the relative availability of schools and classes respectively – are excluded from the profiles because of the high inter-correlations between each of them and indicator 1 (as shown in Table 8.9). These three indicators therefore appear to be measuring similar characteristics and, hence, to include them all would bias the final typology in favour of grouping provinces which were similar on these three indicators only.

The typology is constructed in the same way as that presented in the previous section. The indicators must first be standardized (using Equation 3.3) across all provinces so that those indicators based on index numbers (which have standard deviations of about 10) will not exert too influential an effect on the form of the final typology over those indicators based on either C indices of concentration (which have standard deviations of about 0.05) or the expected years of instruction received or grade attained (which have standard deviations of about 0.6). The standardized values then form the basis for both the unadjusted and adjusted typologies. Both of these typologies are reported in Diagram 10.3.

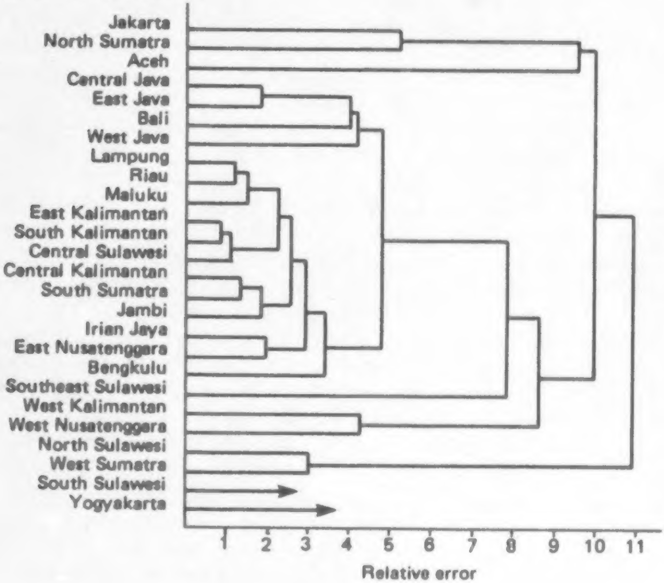
Many of the provinces appear to be very similar in their average profile levels across the six process indicators being used. A point of discontinuity at a relative error value of about 6 can be seen from Diagram 10.3 and hence clustering is stopped there. To that stage 22 of the 26 provinces cluster into four groups. The largest group contains 16 provinces.

The adjusted typology, also reported in Diagram 10.3, indicates Jakarta and North Sumatra have profiles which are sufficiently dissimilar to allow them to remain clustered together in the final typology. This then eliminates one of the clusters from the unadjusted typology. For the same reason, West Kalimantan and West Nusatenggara cannot remain together nor can West Sumatra and North Sulawesi. Bali must also be removed from the Java group. The large group splits into

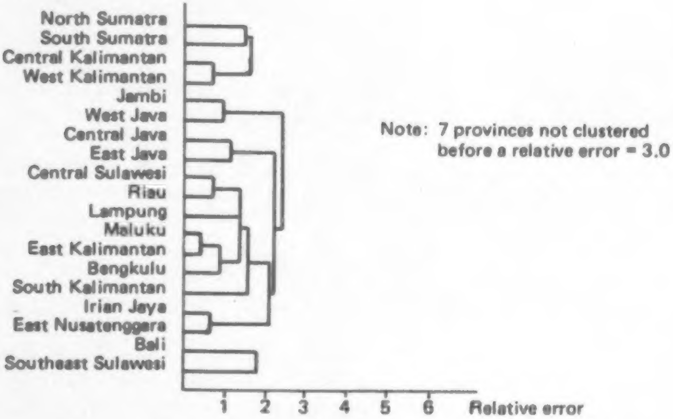
---

[1] A full listing of the data is presented in Johnstone (1978a).

(a) Unadjusted typology



(b) Adjusted typology



Note: 7 provinces not clustered before a relative error = 3.0

Diagram 10.3: *Unadjusted and adjusted typologies across process indicators – Indonesian provinces, 1974* two sub-groups for the final typology with South Sumatra and Central Kalimantan forming a small cluster by themselves. The final typology therefore consists of only three clusters and the average profiles for these across the six indicators are reported in Diagram 10.4. Considerable diversity among the Indonesian provinces is shown by this typology.[1] Also reported in the diagram are the means and standard deviations for each cluster for each indicator. The latter statistic reflects the homogeneity of the clusters. Most values reveal that considerable similarity has been achieved through the clustering algorithm.

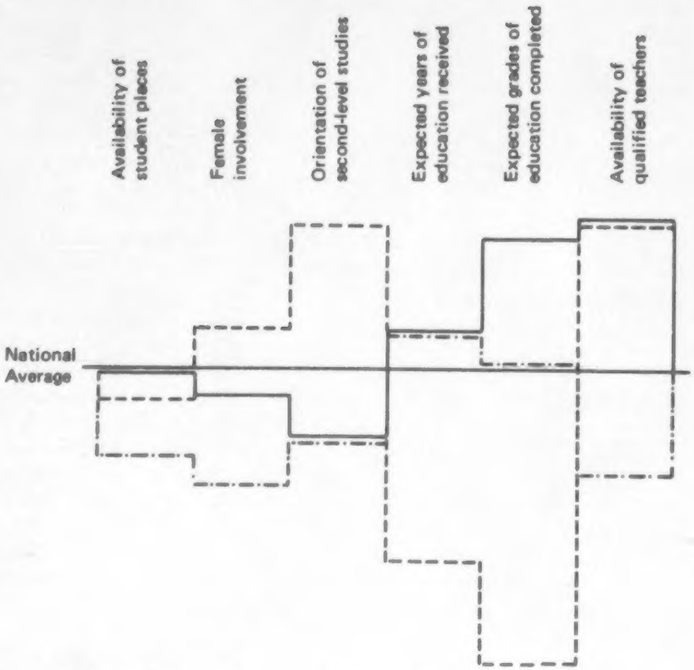
One of the clusters reveals that there is a detectable pattern of education in the island of Java. It contains the three non-city provinces in the island. There is another cluster which contains the least developed of the provinces from the islands of Kalimantan and Sulawesi. The third cluster contains four of the least developed provinces in Sumatra, three of the Eastern Indonesian provinces and three provinces from Kalimantan and Sulawesi. The ten provinces in this cluster would be classified as being in the lower half of all the 26 provinces in their level of economic and general social development. The typology formed here therefore reveals that, except for the three Javanese provinces, no geographical pattern of education system development can be clearly or consistently identified amongst the most economically advanced of the Indonesian provinces. The disparity among all provinces also indicates that any national policy concerned with the process aspects of the provincial systems as measured here would have a different impact in the various areas.

Inspecting the profile for each cluster as drawn in Diagram 10.4 shows the relative strengths and weaknesses of the education systems in each cluster. The two systems in cluster 2, for example, have slightly below national average values for the availability of student places indicator, slightly above average for the female involvement but very low values for the 'expected number of years of education received' and the 'expected number of grades to be completed' indicators. Similarly, although the 'miscellaneous provinces' cluster 3 is distinctly below the other clusters on the availability of student places, the female involvement and the availability of qualified teachers indicators, it is similar to the three (more advanced) Javanese provinces in the other three characteristics.

A more detailed examination of the clusters and of the policies operating would be more appropriate in a report to Indonesian planners and administrators than it would be here. Hence a further in-depth analysis of the patterns is not undertaken here. It must instead suffice to ask that planners in countries other than Indonesia should attempt

---

[1] Diagram 7.2 shows the approximate geographical location of the provinces within Indonesia. The footnote on page 235 also gives a summary description of conditions in the various regions.



Group 1	X̄	99.7	98.3	0.19	5.81	4.84	0.43
	S	2.1	3.8	0.05	0.36	0.33	0.03
Group 2	X̄	98.0	102.0	0.25	5.10	3.70	0.43
	S	1.4	4.2	0.00	0.29	0.35	0.29
Group 3	X̄	94.6	93.4	0.19	5.81	4.47	0.26
	S	4.6	7.1	0.03	0.30	0.28	0.06
Overall	X̄	100.0	100.0	0.21	5.74	4.46	0.33
	S	10.0	10.0	0.05	0.58	0.50	0.12

Note: Membership of the clusters is as follows:

Group 1

Central Java
East Java
West Java

Group 2

Central Kalimantan
South Sulawesi

Group 3

Bengkulu
Jambi
Lampung
Riau
Central Sulawesi
East Kalimantan
South Kalimantan
East Nusatenggara
Maluku
Irian Jaya

Diagram 10.4: *Profiles for three identified clusters of Indonesian provinces – 1974* similar exercises in their own countries. They should then be able to explain how and why particular provinces cluster together. Applying this type of analysis to different years on the same provinces and with the same defined indicators enables the effect of different development growth rates to be detected. The differential impact of national policies in various regions can thus be assessed and evaluated by this means. For the Indonesian example, the analysis could be performed on 1979 data so as to measure the five year differences.

10.5 Résumé

Typology formation has great potential in the contexts of planning and researching education systems development. It forms a starting point from which further work must be undertaken. The effect of policy implementation for example can be assessed once typologies have been formed; different patterns of system development can also be identified in quantitative terms. Typologies offer to research and evaluation relating to education systems development what some of the principles of sampling offer experimental and quasi-experimental designs.

To capitalize on the advantages which can stem from typology formation, it is essential to ensure that profiles across education system characteristics describe proper and general characteristics and not highly specific and/or highly inter-related aspects. Indicators therefore can form a proper basis for the profiles from which a typology can be derived.

In this chapter, it was not appropriate to expand on ways in which typologies can be used in specific planning and research situations. All that could be attempted was a demonstration of how cross-national and intra-national typologies might be formed, and what they would look like. Planners and researchers should now attempt to replicate and extend this presentation in their own situations. It is only through the acceptance of the need for typologies in a number of places that their usefulness will be fully appreciated.

Chapter 11

Implications for the Future

The previous chapters in this book have outlined some of the main concepts for defining, developing and using indicators of education systems. They have also suggested a number of indicators which politicians, planners and researchers might consider using in their examination of education systems. No attempt has been made to cover all aspects of indicators nor to propose a complete listing of all indicators which may be useful. Such undertakings as these would be impossible in the present context because the field of indicators is very much in its infancy. As the field develops further – and the expectation is held that it will – the ideas and suggestions proposed here will be modified and extended. This must happen if indicators are to be adopted and used throughout many education systems in the world. Experience with using indicators in both the political/planning situation as well as in the research situation is thus now the *sine qua non* of further development.

From what has been proposed, five particular important issues relating to indicator development and usage need to be highlighted.

The global nature of the comment made by indicators is important

A first issue – and perhaps the most important if further work is to be supported – is that practitioners who are associated with education systems must seriously consider the basic concepts underlying indicators as defined and incorporate the very different implications of these concepts into their own situations. To date, the emphasis in all forms of measurement and examination of education systems has involved considering very specific features. With the addition of indicators, this emphasis must be changed. Now it is recommended that both specific and broad features be considered as appropriate in particular situations. Sometimes specific features will continue to provide the focus of interest. At other times, it will be preferable to focus on broader features. When this is so, indicators make it easier to obtain an overall view of what an education system is like at a given point in time.

Description is then not confused by the inclusion of particular details and thus the inter-relationships among the general characteristics can be detected fairly readily. In the majority of situations, it is probably sufficient to comprehend and evaluate only these broad patterns. On such occasions, the provision of data pertaining to individual variables only serves to confuse description and prevent the formation of a *gestalt* of what a system is really like. It is therefore recommended that those concerned with the provision and analysis of education systems consider how indicators – and the broad, general descriptions they give – might be incorporated into their own operations.

For some people, confusion might be encountered initially in the adoption of indicators because of difficulties in interpreting the values. For example, educationalists are used to interpreting net enrolment ratios for particular levels in terms of the flow of students into each level of a system. The input indicator measuring total human resource involvement in an education system cannot be so interpreted. It measures the overall flow of students as well as the flow of teachers into the system. It therefore does not comment on what any one level is like. Further it does not comment on the total number of people involved with an education system or on which particular level of the system is biggest or the most developed. It is a more abstract concept and it comments on whether or not a society is allowing a large proportion of its members to be involved with its total formal system of education. The need to interpret indicator values in this broader, more abstract, sense is essential because indicators do not purport to comment on specific features. It is hoped that practitioners accommodate to this change of emphasis and do not abandon the adoption of indicators merely because the frame within which they will have to operate will be at first unfamiliar.

Indicators used must depend on the situational objectives

A second issue is that the indicators used in a particular situation and the way in which their values are interpreted should reflect the objectives which dictate the situation. The earlier discussion noted that for education systems, the identification of these situational objectives is made more complex than it is usually. Indeed the use of indicators in educational planning must accommodate both the objectives of the measurement undertaken as well as the objectives of the particular education system. This distinction is not quite so important in the research application of indicators. In normal measurement operations, it is usual to consider only the objectives of the system and thus make the objectives of the measurement subservient to these. With education systems, there is the additional problem, however, that many objectives are unstated and many outcomes can be not only unplanned but also unexpected. So, use of indicators must depend on not just stated objectives but also on whether the measurement is being conducted for evaluating explicitly stated policies and objectives or for monitoring overall system development or for some other reason. Very often it should be possible to define a suitable set of indicators to monitor the system, and to include in that set a number of indicators explicitly related to the stated system objectives. When this can be done, then the best possible planning use of indicators is being realized.

This overlap between different kinds of objectives dictating the adoption of indicators for education system performance makes it essential that any indicators which are developed should comment in a neutral way and not in a normative way. Progress towards a stated end should not be incorporated into the measurement operation which forms an indicator value. This is not to say that criterion-referenced indicators should never be developed. They can be if the objectives of the measurement operation allow it. Generally, however, it is preferable to define and construct indicators which merely measure the level of a quantity, its distribution or the way in which it has changed in level, and then to interpret the values according to the context of the measurement operation. In some cases, a criterion-referenced interpretation might be what is required; on other occasions, the interpretation might be norm-referenced or self-referenced. The definition of an indicator which in itself is neutral allows this flexibility. It also allows that indicator to be used to monitor and evaluate systems with different objectives and which are governed by different political orientations.

Education systems have many independent characteristics

A third issue which is highlighted by the development of indicators — although it is not really a product of indicators themselves — is that education systems have very many characteristics and most of these characteristics are relatively independent of each other. The major implication which must stem from this observation is that it is not possible to say that a particular system is highly developed overall or that it is a good or a bad system. Each different feature needs individual comment. Certainly some features do measure high involvement or broad expansion and a ranking of systems on these characteristics could be attempted. Indeed this is what is generally done when systems are examined or ordered in research studies. Such an undertaking, however, ignores other equally important characteristics which describe an education system. Hence, although a system may be low on the 'availability of student places' indicator, its values for other indicators could indicate the system has features which are educationally very desirable. To judge a system as being good or bad on just one or two features, therefore, is presently impossible. Perhaps research may, in the future, demonstrate that particular features form crucially important components in a defined hierarchy. Such features could then be used for overall assessment. To date, no such hierarchy has been demonstrated and thus it would seem inappropriate to discuss patterns or norms (such as in the psychological and educational measurement sense) of how education systems should or could develop.

The indicators proposed here are only tentatively suggested

A fourth issue is that, although a large number of education system indicators have been proposed and used in this presentation, these indicators are not intended to form a final inviolate set. As noted above, the stage of the development and use of such indicators is very much akin to that of infancy. The suggestions which have been made here are based on comparatively isolated and unco-ordinated efforts incorporating both empirical and theoretical orientations. Now that a compilation of this previous work is available, it is essential that planners and researchers in particular add to and delete from the lists of indicators proposed. It may also be that some indicators need modification before they are optimally useful in practical situations. If any of these three activities occur, a better understanding of education system development and evolution will also result. Hence, the immediate implementation of the indicators suggested here is advocated. Then both their strengths and weaknesses can be demonstrated in actual planning and research situations. Without the implementation of the proposed indicators, any support, criticism and/or modification will occur mainly on the basis of personal opinion or attitude. What is really required for further development of the indicator sets, however, is more substantial than this.

Not only should change be considered to the actual indicators themselves but ways in which indicators might be used in planning and research should also be re-considered. Some suggestions have been made here but these in no way exhaust the possibilities. One way indicators might be used which is not explored here, for example, is as components in mathematical models. For purposes of explaining and predicting education system development, mathematical models are essential tools. Unfortunately they have not yet realized the potential ascribed to them. Reviews such as those by McNamara (1971) and Johnstone (1974) have shown very clearly that models, as they have been developed to date, neither explain nor predict system behaviour very well or very rationally. Both of the above reviews also recommend that careful consideration be paid to identifying what components should be incorporated into the models. By examining broad rather than specific features of systems, indicators might prove to be components which allow planners and researchers to capitalize on the potential power and capability of mathematical models. If this did happen, policies would be able to be formed more carefully and with a greater understanding of the reasons for their limitations and for the potential of extending them at a later date. This in turn might lead to increased rationality in debate between and within political parties over the importance of different emphases in various political parties.

There are other ways in which indicators can and should be used. Efforts must be made to encourage planners and researchers to explore the potential of these alternatives.

Multiple bases for indicator development are available

A fifth issue is that indicators have many statistical bases from which they can be formed. The use of these different bases to form indicators defined within a single set in no way complicates the interpretability of the indicators. It does, however, allow a greater variety of characteristics to be measured, and this is a major advantage of indicators themselves. To date, the definition of variables has very much concentrated on total quantities, and it thus has ignored the important considerations of flows and distributions.

In the indicator sets proposed, a number of different statistical bases have been used. Not all possibilities are explored, however. In Chapters 3 and 4, for example, a number of other bases are also discussed. These additional possible bases should serve as a guide for the development of indicators to measure other purposes and objectives. Such indicators as might arise can then be generalized into other situations and thus be included in a basic set of indicators suitable for evaluating and monitoring all similar types of systems. If this were to happen, interpretability would not be reduced but the completeness of comment on education system development would be increased.

To those who will undertake the definition and development of additional indicators, a cautionary note must be issued at this point. It is that a balance must be maintained between the application of statistical principles and the interpretability, especially for laymen, of the indicators developed. In some of the indicators presented here, for example, the results of complex statistical techniques have been modified so that simple arithmetic is finally used for the calculation of actual indicator values. This alteration, although not changing the importance of the empirical result, helps to ensure discussion is not avoided on the grounds that the mathematics of the final indicator form are too complicated. There is no point in defining an extremely useful or important indicator if it takes an advanced statistican to interpret it and to appreciate the nuances of the comment it makes. On the other hand, people who are untrained in any form of simple arithmetic or who consider they have an aversion to anything statistical or mathematical, must try to overcome their 'disability' and/or prejudice and understand at least the concept being measured by an indicator if not the actual mathematics involved. The basis for an indicator must be statistical to some extent – this cannot be circumvented. It is therefore essential that a balance between statistical principles and interpretability be sought. This is perhaps especially so in the early stages of adopting indicators because if the concepts of indicators and their statistical bases are not accepted immediately, there is little chance of their widespread adoption in the long term as planning and policy-making tools.

Finale

From what has been said, it should be obvious that the ideas which motivated this book evolved from the firm belief that proper planning and researching of education systems must be based on sound quantitative methods. To date, we have certain methodologies but, as planners and researchers, we are not sure about the characteristics to which we should apply our methods. Indicators provide a possible solution to that problem. To adopt them, however, requires that many advances be made. These advances include an acceptance of the concept of global rather than specific measurement and the revision and extension of present methods of data collection and reporting. Given that both the educational and statistical concepts underlying this approach to measuring the development level of education systems are acceptable to a variety of politicians, planners, administrators and researchers, a more intelligible complete and predictable explanation of the evolution of these systems should be forthcoming.

Appendix A

The Calculation of Values for the Gini Coefficient

As is indicated in Chapter 3, the Gini coefficient provides a statistical basis for measuring the degree to which a particular characteristic is equally distributed throughout an education system. The way in which the degree of equality is ascertained involves relating the distribution of the characteristic being considered to the distribution of some appropriate criterion variable.

The selection of an appropriate criterion is an important consideration of using the Gini coefficient partly because of the nature of the coefficient itself. As was shown in Section 3.3, the coefficient is one of a number of possible distributive indicators measuring the dispersion of a characteristic throughout an education system. As it is not invariant of the units of measurement used, however, the Gini coefficient must relate the cumulative percentage (or proportion) of the characteristic to the cumulative percentage (or proportion) of the criterion if a fully interpretable and generalizable value is to be calculated. Conrad (1954) and Kendall and Stuart (1958, 46-49) discuss the need to calculate the coefficient in this way if meaningful results are to be obtained which are comparable from one situation to another. If the calculation is made without obtaining ratio values of some sort, the value for the Gini coefficient will reflect the unit in which the characteristic is measured (be they people, schools, monetary units, etc).

One equation for calculating the Gini coefficient has already been reported. It is:

$$G = \sum_{i=1}^{n} (p_{i-1} q_i - p_i q_{i-1}) \qquad \textit{(Equation A.1)}$$

where p_i = cumulative proportion of the characteristic whose equality is being investigated

q_i = cumulative proportion of the variable which is acting as the criterion for the measurement.

The value for both p_0 and q_0 is 0.0 while the value for both p_n and q_n is 1.0. If percentages rather than proportions are used, these limiting values change to 0.00 and 100.0 respectively while the values calculated by the above equation should be divided by 10,000.

There are six steps to follow in the calculation of values:

Step 1
Nominate the characteristic for which the degree of equality is to be assessed as well as the criterion which is to provide the standard for comparing the degree of equality. The criterion should be related in some way to the characteristic. Generally it will measure some aspect of the population distribution but this does not always have to be the case. Sometimes, for example, the criterion might be another variable measuring a different aspect of education system provision. Both the characteristic and the criterion will be measured across appropriate units such as provinces, states, geographical areas, etc.

Step 2
Calculate the proportion of the total amount (or distribution) taken by each unit for the characteristic and for the criterion separately.

Step 3
Rearrange the units on which the proportions have been calculated in ascending order of the proportions for the characteristic.

Step 4
Accumulate the proportions for both the characteristic and the criterion separately in the order of the units established in step 3. The final value for each distribution should therefore be 1.0.

Step 5
The proportions calculated can now be plotted to form the Lorenz curve. The line of equality is drawn from the point (0.0, 0.0) to the point (1.0, 1.0).

Step 6
Substitute in the formula as stated above.

EXAMPLE
An example demonstrates the way in which these steps are followed when the coefficient value is being calculated. The data used for this example are those reported in Table 3.7 for the distribution of enrolments across the six grades at the first-level in El Salvador in 1965. These data are reported again in Table A.1.

Grade	Proportion of total enrolments (characteristic) Y_i	Proportion of total education (criterion)	Cumulative proportion of enrolments p_i	Cumulative proportion of education q_i
1	0.37	0.167	0.37	0.17
2	0.20	0.167	0.57	0.33
3	0.15	0.167	0.72	0.50
4	0.12	0.167	0.84	0.67
5	0.09	0.167	0.93	0.83
6	0.07	0.167	1.00	1.00

Table A.1: *Sample calculation for the Gini coefficient – data for El Salvador from Table 3.7 – distribution of first-level enrolments by grade*

Step 1

The characteristic here is the proportion of enrolments in each grade while the criterion is the amount of education being offered in each grade as a proportion of the total first-level education offering; that is, the distribution of education by grade.

Step 2

The proportions for the characteristic and the criterion are reported in the first two columns of Table A.1. Those for the characteristic have been taken from the Unesco *Statistical Yearbook* 1974 while those for the criterion are merely calculated by assuming each grade in effect offers one-sixth of the total first-level education (because there are six grades).

Step 3

Table A.1 reports the units (in this case the grades) ranked in ascending order on the characteristic (ie the proportion of the total enrolment accounted for by each grade).

Step 4

Columns 4 and 5 of Table A.1 report the accumulation of the proportions calculated for each grade. Both cumulative proportions for grade 6 equal 1.0 as should be the case.

Step 5

Diagram 3.2 shows the Lorenz curve and the line of equality. In this case, it is more convenient to plot the characteristic along the horizontal axis and the criterion along the vertical axis. Often the reverse of this situation will be the case. It does not really matter, however, because neither the value of the Gini coefficient nor its interpretation are altered.

Step 6
The values are now substituted into the equation as follows:

$$\begin{aligned} G = {} & (0.0 \times 0.17 - 0.37 \times 0.00) + (0.37 \times 0.33 - 0.57 \times 0.17) \\ & + (0.57 \times 0.50 - 0.72 \times 0.33) + (0.72 \times 0.67 - 0.84 \times 0.50) \\ & + (0.84 \times 0.83 - 0.93 \times 0.67) + (0.93 \times 1.00 - 1.00 \times 0.83) \\ = {} & (0.0 - 0.0) + (0.1221 - 0.0969) + (0.2850 - 0.2376) \\ & + (0.4824 - 0.4200) + (0.6972 - 0.6231) + (0.93 - 0.83) \\ = {} & 0.31 \end{aligned}$$

This is the value which is reported in Section 3.3.

Another formula for computing the Gini coefficient is proposed by Sen (1973, 31). It does not require adding the ranked proportions together as for step 4. All that is required is for the units to be ranked in descending order of the proportions for the characteristic (ie up to step 3). Sen's formula is then:

$$G = 1 + \frac{1}{n} - \frac{2}{n^2 \bar{x}}[y_1 + 2y_2 + - - + ny_n] \qquad \textit{(Equation A.2)}$$

where n = number of units
$\bar{x}$ = average value of characteristic across all units
y_i = proportion of characteristic in unit i.

This formula holds provided the proportions of the characteristic have been ranked in descending order, ie:

$$y_1 \geqslant y_2 \geqslant y_3 \geqslant \ldots \geqslant y_n$$

Steps 1–3 in the above list are the same for this formula. After that:

Step 4 Calculate the average proportion and note the total number of units.
Step 5 Substitute in formula A.2.

EXAMPLE

The results of steps 1–3 are reported in Table A.1 in columns 1 and 2.

Step 4 $n = 6$

$$\bar{x} = \frac{1.00}{6} = 0.167$$

Step 5

$$\begin{aligned} G = {} & 1 + \frac{1}{6} \\ & - \frac{2}{36(0.167)}[0.37 + 2(0.20) + 3(0.15) + 4(0.12) + 5(0.09) \\ & \qquad + 6(0.07)] \\ = {} & 1 + \frac{1}{6} - \frac{2}{36(0.167)}[2.57] \\ = {} & 1.167 - 0.855 \\ = {} & 0.31 \end{aligned}$$

Both formulae yield the same value naturally and so it is a matter of personal choice which is used. The former provides the values to be used in plotting the Lorenz curve while the latter is probably a little simpler computationally.

When using education system variables, distributions of the characteristic and/or of the criterion can sometimes be distorted from what is, from a statistical point of view, most desirable. This distortion can be manifest in two different situations.

Firstly, it could be that because of the way in which a situation was defined, the proportions of the criterion exceed the proportions of the characteristic for all units of the distributions. This situation does not raise any problem, however. Substituting into the equation results in a negative value. Ignoring the sign and simply taking the absolute value produces the same value for the coefficient as would have resulted had the criterion and the characteristic been reversed.

Secondly, the proportions of the criterion could exceed the proportions of the characteristics for some of the intermediate units. The result of this is that the Lorenz curve can cross the line of equality and this results in a distorted calculated value for the Gini coefficient. When this does happen, the extent of the inequality must be maximized. To do this, the characteristic should be ranked (as usual) in ascending order but, when overlap occurs, units should be ranked in descending order of the criterion. This will generally ensure that the inequality is maximized and that the Lorenz curve falls below the line of equality for all units.

Appendix B

Example of Indicator Development by the Method of Factor Analysis

As is described in Chapter 4, factor analysis is perhaps the main way to develop indicators by empirical methods. One of the purposes of this Appendix is to demonstrate the way in which a set of education systems data may be analysed so as to identify and further develop possible indicators. The presentation is thus reported as a normal factor analysis should be.

A second purpose in reporting the analyses here is to demonstrate the cross-temporal stability of a set of previous results (Johnstone, 1977). Such a demonstration is essential if basic indicators are to be truly useful for planning and research purposes. McGranahan *et al* (1972, 15), Adelman and Morris (1967, 144) and Peaslee (1969), all warn that cross-national patterns are not necessarily stable cross-temporally. If shifts in the importance of characteristics measured by indicators occur from one period to another or variables differ substantially in their usefulness in defining the indicators, then neither the indicators nor the variables can be validly or reliably used to measure, in a norm-referenced way especially, educational growth across systems. They would then be indices of the type Anderson refers to as 'rubber rulers' (1961, 10) and not indicators Coombs demands should be developed for 'assessing the performance of education systems and their various parts' (1969, 34).

Any replication analysis in a cross-national situation must take the form of analysing, using the same techniques, the same variable set across basically the same set of countries. The data must relate to another year. The year selected should be such that a sufficient time has elapsed from the date of the original analysis. This allows the data to be less contaminated by influences which may have affected development for or around the year originally analysed.

The original report analyses data related to 1969. A set of 89 input and process variables was used. The year 1960 is selected for the replication analysis. This year provides a sufficient time period before 1969 to allow proper comparisons to be made. When data for, say, 1979 or 1980 become available, the analysis could be replicated again. The selection of 1960 in no way implies that conditions in that year were necessarily the same as they were in 1969. The earlier year simply allows the required time interval. It also allows educational progress to be mapped during the decade often referred to as 'the education decade'.

A replication analysis of the input set is reported in Section 4.1. This appendix therefore reports only the analysis of the process variables. The methodology employed is the same as that in the previous report:

- ☐ the method of principal components factor analysis is used
- ☐ the computer program is BMDO3M
- ☐ communality estimates are squared multiple correlations
- ☐ factors are extracted for rotation on the basis of their eigen values being greater than 1.0
- ☐ Varimax rotation is performed.

The combination of all these features aims to identify those basic components of a variable set measuring the process aspects of education systems which best facilitate differentiation among a set of entities. The variables so identified should also form the most parsimonious basic set. The variables analysed (and their acronyms) are reported in Table B.1 while the factors are summarized in Table B.2. The countries used are reported in Table B.3 together with their factor scores.

The first factor describes 'the extent and refinement of basic educational provision' within a country. A country having high scores on this factor has developed its second-level system with few, if any students in the teacher-training stream. There are high promotion and retention rates in the early years of the first-level. The building programme is extensive and an attempt has been made to broaden the range of educational activities by the development of special education facilities.

The second factor reflects the 'extent of female involvement in education'. All variables related to the percentage of either pupils or teachers who are female have high positive loadings. The factor thus indicates those countries having relatively high percentages of females participating at all levels of education. This in no way implies equality of provision in a country having high factor scores.

The third factor complements the first by describing 'the development emphasis on second-level education'. The rate of change in second-level enrolments compared to those at the first-level is high and, probably because of this, the rate of change for the ratio of third- to second-level enrolments is low. Such a situation could only happen when second-level enrolments are increasing rapidly. The increase appears to be centred in both the general and technical education streams especially for males.

The fourth factor describes the major study orientation within the third-level institutions of a country. It clearly differentiates countries whose third-level student enrolments emphasize law/social science courses from countries whose student enrolments emphasize humanities/education/fine arts courses.

Variable	Interpretation of Variable	Variable Code
1. Ratio of male enrolments at the pre-school level to male enrolments at the first-level	Indicates relative size of vertical divisions on separate sides of the educational pyramid of a country	P/1M
2. Ratio of male enrolments at the second-level to male enrolments at the first-level		2/1M
3. Ratio of male enrolments at the third-level to male enrolments at the second-level		3/2M
4. Ratio of female enrolments at the pre-school level to female enrolments at the first-level		P/1F
5. Ratio of female enrolments at the second-level to female enrolments at the first-level		2/1F
6. Ratio of female enrolments at the third-level to female enrolments at the second-level		3/2F
7. Ratio of enrolments in special education to enrolments at the first-level	Indicates extent of provision of special education facilities	SPEC/1
8. % of first-level enrolment who are female	Indicates relative size of horizontal divisions on each side of the educational pyramid of a country	% F-P1
9. % of second-level enrolment who are female		% F-P2
10. % of third level enrolment who are female		% F-P3
11. Proportion of all second-level male enrolments in the general education stream	Reflects relative emphasis between streams at the second-level of education within a country	% 2 M GN
12. Proportion of all second-level female enrolments in the general education stream		% 2 F GN
13. Proportion of all second-level male enrolments in the teacher training stream		% 2 M TT
14. Proportion of all second-level female enrolments in the teacher training stream		% 2 F TT
15. Proportion of all third-level male enrolments following humanities/education/fine arts courses	Reflects relative emphasis between streams at the third-level of education within a country	% 3 M HEF

Variable	Interpretation of Variable	Variable Code
16. Proportion of all third-level female enrolments following humanities/education/fine arts courses	Reflects relative emphasis between streams at the third-level of education within a country	% 3 F HEF
17. Proportion of all third-level male enrolments following law/social science courses		% 3 M LSS
18. Proportion of all third-level female enrolments following law/social science courses		% 3 F LSS
19. Proportion of all third-level male enrolments following natural science courses		% 3 M NS
20. Proportion of all third-level female enrolments following natural science courses		% 3 F NS
21. Number of student places provided at the third-level per student studying at the third-level in another country	Reflects extent to which social demand for third-level education is met by a country's own educational provision	HOME/ABD 3
22. Proportion of total first-level enrolment in first two grades compared to expected	Reflects distribution of enrolments across grades and effects of early dropout and repetition	ACT/EXP 1
23-29. Rate of change in variables 1-7	Indicates relative emphasis in expanding enrolments at different levels of the educational pyramid in a country	ROC precedes the appropriate variable code listed above
30-32. Rate of change in variables 8-10	Indicates rate at which provision for females (and thus indirectly for males) is changing	
33-42. Rate of change in variables 11-20	Indicates changing emphasis of enrolments in the various streams at second- and third- levels	
43. Rate of change in variable 21	Reflects attempt at accommodating the demand for third-level education within a country	
44. Rate of change in variable 22	Reflects changing distribution of enrolments at the first level	

Variable	Interpretation of Variable	Variable Code
45. Proportion of teachers at the first-level who are female	Indicates the relative responsibility between the sexes for educational instruction	% F-T1
46. Proportion of teachers at the second-level who are female		% F-T2
47. Pupil-teacher ratio at first-level	Reflects relative provision of teaching personnel per pupil enrolled	P-T1
48. Pupil-teacher ratio at second-level		P-T2
49. Pupil-teacher ratio at third-level		P-T3
50-51. Rate of change in variables 45-46	Indicates relative change in responsibility for instruction between the sexes of the teachers	ROC precedes the variable code listed above
52-54. Rate of change in variables 47-49	Reflects difference in rates at which teachers can be trained and/or recruited compared to social demand for a particular form of education	
55. % of total recurrent expenditure spent at the first-level	Indicates relative part of total recurrent budget taken by each level	% $ 1
56. % of total recurrent expenditure spent at the second-level		% $ 2
57. % of total recurrent expenditure spent at the third-level		% $ 3
58. % of total finance budget spent on capital expenditure	Indicates relative importance of building programme	% $ CAP
59. % of total recurrent expenditure spent on teacher salaries	Indicates proportion of recurrent budget already committed for non-developmental activity	% $ T SAL
60-62. Rate of change in variables 55-57	Indicates change in financial provision and thus relative emphasis placed between levels	ROC precedes the variable code listed above
63. Rate of change in variable 58	Indicates change in emphasis on building programme over time	

Variable	Interpretation of Variable	Variable Code
64. Rate of change in variable 59	Indicates change in proportion of recurrent budget which can be used for development purposes	ROC precedes the variable code listed above
65. Age of starting first-level	Indicates main structural elements of educational system provision	AGE START
66. Formal length of first-level		LENGTH 1ST
67. Formal length of second-level general education stream		LENGTH 2ND
68. Maximum number of years of compulsory education		YRS COMP
69. Minimum number of years of general education prior to entering teacher training	Indicates relative level of education and training presently received by teachers	YRS BEF TT
70. Level at which majority of teacher training is conducted		LEV MAJ TT
71. Practice underlying certification at end of first-level	Indicates certification practice at first-level	EXAMS 1ST

Table B.1: *The process variables – their interpretations and codes*

The fifth factor in some ways complements the second. It has a negative loading on the 'percentage of first-level students who are female' variable but high positive loadings for the rates of change in both this variable and the corresponding one for the second-level. This combination is to be expected because, as equality is reached, the rates of change must approach zero. The increase in the proportion of third-level male students in law/social science courses appears to be an attendant yet presumably educationally independent feature in countries where such changes occur. The main characteristics of the factor identify it as reflecting a country's 'development emphasis on providing an equality of female pupil involvement'.

The sixth, eighth and tenth factors all identify particular aspects of an education system which distinguish between the educational provision of various countries. These aspects are:

☐ the proportion of second-level enrolments in the general education stream

☐ the extent of pre-school provision compared to that at

Factor	Variable	Loading
1	2/1 M	0.89
	2/1 F	0.85
	SPEC/1	0.65
	% 2 M T T	–0.52
	% 2 F T T	–0.45
	ACT/EXP 1	–0.55
	% $ CAP	0.48
	LENGTH OF 2ND	0.59
2	% F-P 1	0.55
	% F-P 2	0.73
	% F-P 3	0.80
	% F-T 1	0.78
	% F-T 2	0.88
3	ROC 2/1 M	0.83
	ROC 3/2 M	–0.87
	ROC 2/1 F	0.84
	ROC 3/2 F	–0.57
	ROC % 2 M T T	–0.71
4	% 3 M HEF	–0.80
	% 3 F HEF	–0.74
	% 3 M LSS	0.80
	% 3 F LSS	0.86
5	% F-P 1	–0.51
	ROC % F-P 1	0.79
	ROC % F-P 2	0.66
	ROC % 3 M LSS	0.85
6	% 2 M GN	0.91
	% 2 F GN	0.89
7	% 2 F T T	0.48
	ROC % 3 M HEF	0.83
	ROC % 3 F HEF	0.88
	P-T 2	–0.57
8	PRE/1 M	0.93
	PRE/1 F	0.95
9	ACT/EXP 1	–0.62
	YRS BEF T T	0.85
10	% 3 M N S	0.71
	% 3 F N S	0.85
11	HOME/ABD 3	0.85
	% $ 1	–0.47
12	P-T 1	0.58

Factor	% of total variance	% of common variance
1	11.5	13.6
2	10.0	11.9
3	9.2	10.9
4	8.5	10.1
5	7.9	9.4
6	6.9	8.2
7	6.7	7.9
8	6.4	7.6
9	5.3	6.3
10	5.3	6.3
11	3.8	4.5
12	2.8	3.3
Total	84.3	100.0

Table B.2: *Summary of factor loadings of variables defining the 1960 process analysis factors*

Country	Factor number									
	1	2	3	4	5	6	7	8	9	10
Albania	−1.57	−2.05	0.41	−0.84	−1.00	−2.15	−0.36	−0.11	−0.25	0.27
Argentina	−0.40	1.09	0.00	0.91	0.38	−1.88	1.23	−0.40	−0.18	−0.65
Australia	0.58	0.42	−0.13	−1.30	−0.16	0.73	0.33	−0.45	−0.31	−0.12
Austria	2.02	−0.54	−1.34	0.75	0.16	−1.76	−0.17	0.64	−0.29	−0.03
Brazil	0.01	0.57	−0.03	0.69	0.22	0.69	0.26	−0.26	−0.54	0.00
Canada	−0.06	0.38	−0.13	−0.69	0.22	0.51	0.27	−0.54	5.45	0.17
Chile	−0.06	0.59	−0.24	−0.29	−0.27	0.49	0.03	−0.63	−0.08	−0.84
Columbia	−0.48	−0.14	0.32	0.29	−0.09	−0.30	1.76	−0.62	−0.28	0.98
Costa Rica	−0.92	0.81	−1.20	−2.02	−1.07	0.76	0.10	−0.92	−1.24	−0.79
Cuba	−1.08	1.61	0.54	0.69	−0.27	0.92	−0.36	1.08	−0.75	−0.10
Denmark	1.05	−0.58	−0.73	−0.48	−0.39	−0.11	1.39	−0.24	1.00	−0.87
El Salvador	−0.34	−0.64	−0.75	−0.02	−0.67	−0.39	2.59	−0.25	−0.90	1.33
Ethiopia	−0.75	−1.12	0.04	−0.81	5.89	−0.38	−0.68	−0.96	−0.88	0.80
Finland	1.31	1.16	1.91	0.03	0.12	0.65	−0.54	−0.47	−0.61	−0.56
France	0.04	0.31	0.79	0.67	−0.26	1.04	−0.76	1.41	1.07	3.90
Germany, Fed. Rep.	2.47	−0.86	−1.50	−0.18	0.16	−1.14	0.44	0.98	0.32	−0.11
Guatemala	−0.75	−0.28	−0.37	1.41	−0.03	0.44	0.85	0.56	−0.77	−0.70
Haiti	0.26	−0.73	−0.16	1.81	0.95	0.61	0.46	−0.99	1.45	−2.11
Honduras	−0.65	0.04	0.13	1.40	−0.48	−0.16	−0.26	−0.61	−0.37	−0.16
Hungary	−0.44	0.58	3.89	−0.16	0.09	−1.48	−0.16	0.86	0.83	−1.14
India	−1.33	−2.09	−0.22	−2.38	−0.21	1.86	0.04	2.68	−0.54	−1.42
Iran	−0.42	−1.45	−0.58	−0.08	0.41	1.35	0.06	0.04	0.62	−0.72
Iraq	−0.87	−1.33	−0.37	−0.61	−0.17	0.93	−1.96	−0.12	−0.02	1.22
Ireland	0.14	1.14	0.29	−1.07	−0.08	0.16	−0.43	−1.55	−0.16	−0.10
Israel	−0.42	0.61	0.21	−0.57	0.15	−0.49	−0.28	1.78	0.39	0.17

Country	Factor number									
	1	2	3	4	5	6	7	8	9	10
Italy	1.42	0.42	−0.01	0.78	0.86	0.79	0.98	2.07	−0.67	−0.06
Japan	1.62	−1.06	−0.12	0.01	−0.70	0.71	−0.46	−0.17	0.24	−1.10
Korea, Rep. of	0.23	−1.84	−0.64	0.48	−0.08	0.58	−0.88	−0.83	0.45	0.00
Mexico	−1.06	−0.40	1.12	1.33	−0.50	0.35	−0.36	0.10	−0.39	0.31
Netherlands	1.76	−0.58	0.27	0.48	−0.08	−0.94	−0.43	1.77	−0.34	0.33
New Zealand	0.06	0.40	−0.26	−1.20	−0.63	0.50	−0.38	−0.14	−0.25	−0.03
Norway	0.98	−0.76	0.60	−1.33	−1.02	−0.38	0.31	−1.83	−0.25	1.21
Panama	−0.05	1.76	0.06	0.10	0.12	0.70	−0.48	−0.60	−0.69	0.59
Paraguay	−1.04	0.42	0.32	0.13	0.51	0.49	3.58	−0.31	0.12	0.68
Peru	−0.37	−0.20	0.34	0.12	0.21	0.53	−0.11	−0.24	−0.13	−1.23
Poland	−0.41	0.79	−0.04	−0.03	0.22	−1.79	−0.25	0.32	0.19	−0.16
Portugal	−0.11	−0.44	0.09	−0.47	−0.42	−0.63	−0.29	−1.21	0.29	2.14
Romania	−1.14	−0.02	0.28	−0.53	−0.86	−1.41	−0.01	1.53	0.38	0.26
Spain	0.25	−0.48	−0.04	−0.15	−0.10	0.19	0.39	1.13	−0.46	0.81
Sweden	2.25	0.61	1.63	−0.56	0.36	−0.31	−0.52	0.05	−0.73	−0.50
Thailand	−0.14	0.13	−1.89	2.62	−0.15	1.23	−0.27	0.03	−0.07	0.35
Turkey	−0.54	−1.14	0.69	1.50	−0.78	0.30	−1.35	−1.00	−0.28	−0.94
United Kingdom	1.66	0.74	−0.25	−1.19	−0.36	0.85	−0.73	−1.25	−0.74	0.80
USA	−0.34	1.66	−0.30	0.45	0.01	1.12	−0.82	0.37	0.61	−0.04
USSR	−1.08	2.38	−2.80	−1.10	0.76	−0.89	−1.01	1.06	0.54	−0.60
Venezuela	−0.46	0.10	0.12	0.45	−0.31	−0.50	1.10	−0.99	−0.52	−0.80
Yugoslavia	−0.85	0.00	−1.13	0.46	−0.68	−2.37	−1.85	−0.80	−0.07	−0.45

Table B.3: *Factor scores for each country on the 1960 process dimensions*

the first-level
- ☐ the proportion of third-level students enrolled in natural science courses.

These three independent dimensions thus indicate specific features within a total systems provision which allow countries to be reliably and validly compared and contrasted.

The seventh factor has high factor loadings on four variables directly or indirectly associated with teacher training. Countries with high factor scores here are increasing the proportion of their enrolments in the third-level humanities/education/fine arts courses. They also have a low second-level pupil-teacher ratio – possibly lowered by the inflow of recent third-level education graduates into the second-level teaching force. A less highly loaded variable is the proportion of females in the second-level teacher training strand. This form of training is significant in those countries which have not yet developed their third-level facilities fully and which still need to depend on this form of training. This factor thus reflects the 'emphasis on quantitatively expanding teacher education facilities'.

The ninth factor describes 'the general quality of educational provision'. It is highly loaded on two variables which can be directly interpreted from a qualitative viewpoint. Countries with high factor scores here have low repetition and drop-out rates in the first-level and their teacher trainees receive a comparatively long period of general education before they enter their training programmes.

Factors 11 and 12 each account for an insignificant part of the total variance. Their factor loadings also indicate unique variance might now be entering the factor space. Hence, although they can be interpreted, neither factor is included in the set of basic dimensions.

The ten factors finally interpreted account for 77.3 per cent of the total variance of the defined space.

Comparison between the 1960 and 1969 process analyses

A number of important observations can be made by comparing the 1960 and the 1969 process dimensions of education systems.

The first to emerge from the cross-temporal comparisons is that there are five aspects of educational provision which remain stable in their ability to discriminate among education systems. These aspects are:

- ☐ the relative emphasis placed by third-level students on humanities/education/fine arts courses or law/social science courses
- ☐ the extent and refinement of basic educational provision
- ☐ the extent of female involvement amongst pupils and teachers
- ☐ the emphasis placed on second-level general education

☐ the emphasis placed on pre-school education.

The first three of these aspects form into separate factors in each of the analyses; the remaining two form into separate factors in the 1960 analysis and into opposite poles of a bipolar factor in the 1969 analysis. Hence these are five aspects which can be used either independently or collectively to compare and contrast education systems and their development through at least the 1960-69 period. They can therefore be regarded as indicators in their own right for monitoring systems developments over this period. Their ability to form such a basis is allowed by their identification in both years of analysis. They thus identify areas both planners and researchers should focus upon very critically.

The second observation is that there has been a major shift in emphasis within education systems from the second- to the third-level. This accords with Peaslee's (1969) hypothesis that the level of development emphasis within a system moves slowly upwards. In the 1960 dimensions, the most important factor concerned educational provision mainly at the first- and second-levels. In the 1969 dimensions, this factor is again extracted but it is less important, in terms of amount of variance explained, than is the extensiveness of total provision dimension which includes variables related to third-level education. The shift also shows up in the replacement of the 1960 development emphasis on second-level education factor by the 1969 development emphasis on higher education factor. Both of these changes show that the 1960 decade witnessed an increasing emphasis on extending provision into the third-level of education systems.

This shift also has certain other attendant consequences and these are manifest in some of the other factors extracted for the two years. For example, the increase in third-level provision causes more teacher training to be offered at this level. Hence, the 1960 quantitative development of teacher education dimension merges into the 1969 major study orientation within the third-level dimension. This causes the latter to be interpreted more specifically rather than across the broad humanities/education/fine arts spectrum of courses.

Summary

The factors identified in these analyses can each be regarded as providing a basis for indicator development. They can be used as they are formed. When they are, the factor scores reported in Tables 4.2 and B.3 provide values for each factor which can be incorporated directly into statistical analyses. This approach is the one adopted in part in Chapter 9. The identified factors can also be regarded as forming a basis from which the significant variables defining each factor can be isolated. These can then be used to form purer indicator scores. Such an approach as this is partially adopted in Chapter 5. Both of these methods of utilizing the results of the factor analysis are perfectly acceptable in the definition of education system indicators. That they are reflects the value of using this particular multivariate technique at least as a first step in the empirical definition of indicators.

Appendix C

Technical Details of Forming Typologies

In this technical Appendix, a summary is presented of the way in which clusters such as those described in Chapter 10 can be formed. As noted earlier, many methods can be used to form clusters. The method described here uses the Euclidean distance measure to estimate the similarity between the profiles for ten education systems and then the diameter method to determine which pairs of profiles are most similar. The same technique can be used to form both the unadjusted and the adjusted typologies.

The first step in the process is to calculate the values for each indicator which is to form part of the profile of the education systems. In Section 10.3, four of the input indicators are used and hence the profile is formed across these. The values for the component variables of each index are standardized across the 85 nations listed in Diagram 10.1 and then combined in the way described by Equation 3.4. Equal (and unit) weights are used because no variable is considered as being more influential than any other. The sums of the standardized scores for each indicator are then divided by the number of variables used so as to ensure the distributions of the four indicators have equal standard deviations. This equality is important in typology formation because it ensures no one indicator is more dominant than another in the determination of membership of a particular cluster.

The indicator values for a sample of ten of the 85 countries are reported in Table C.1. The value for indicator 3 for Afghanistan, for example, were calculated as:

$$I = \frac{\dfrac{1.5 - 4.075}{1.692} + \dfrac{10.0 - 16.239}{6.099}}{2}$$

where 4.075 and 1.692 = the mean and standard deviation respectively for the % of GNP spent on education and calculated across all 85 nations being considered

16.239 and 6.099 = the mean and standard deviation respectively for the % of government

312

expenditure spent on education and calculated across all 85 nations being considered

1.5 and 10.0 = values for Afghanistan for 1969 as reported in the Unesco *Statistical Yearbook*.

Country	Involvement of human resources	Development emphasis on changing human resources	Financial involvement	Financial expansion
Afghanistan	−1.799	1.229	−1.272	−0.292
Burma	−0.461	1.223	−0.206	0.182
Cameroon	−0.621	0.549	0.035	0.970
Central African Rep.	−1.331	0.339	−0.268	0.343
Ethiopia	−1.874	1.874	−1.090	0.429
Iran	−0.556	−0.658	−1.091	0.080
Liberia	−0.903	0.991	−0.838	−0.248
Niger	−2.203	1.073	−0.792	0.364
Senegal	−1.314	0.639	−0.063	0.200
Somalia	−1.958	1.036	−1.295	−0.097

Note:
Indicator values have been calculated from variables (defined in Table 5.1) and standardized across the 85 countries listed in Diagram 10.1.

Table C.1: *Values for four input indicators calculated for a sample of ten countries*

From Table C.1, the D^2 values for profile similarity can be calculated using Equation 10.1. For Afghanistan and Burma, for example,

$$
\begin{aligned}
D^2 &= (-1.799 + 0.461)^2 + (1.229 - 1.223)^2 + (-1.272 + 0.206)^2 \\
&\qquad + (-0.292 - 0.182)^2 \\
&= 3.151
\end{aligned}
$$

The matrix for all 45 D^2 values thus calculated is presented in Table C.2. These values may now be used to form the unadjusted typology. If the adjusted typology was to be formed, Equation 10.2 rather than Equation 10.1 would have to be used.

Convenient ways to begin the clustering are to look for the lowest of all D^2 values in the matrix or to select the first country in the matrix. Both of these approaches lead to the same result here – that is a link between Afghanistan and Somalia at $D^2 = 0.101$. This then is the first pair to form.

Remaining with this pair of countries, the next smallest value for Afghanistan is $D^2 = 0.848$ for Niger. The corresponding value for Somalia is $D^2 = 0.527$. If Niger is to join this pair therefore, it must not be able to link with any other profile at a D^2 value less than the maximum of these two values – that is 0.848. It can, however, link

	Afghanistan	Burma	Cameroon	Central African Republic	Ethiopia	Iran	Liberia	Niger	Senegal	Somalia
Afghanistan	–	3.151	5.151	2.422	0.975	7.593	1.049	0.848	2.287	0.101
Burma	3.151	–	1.159	1.568	3.263	1.122	0.834	3.434	1.089	3.540
Cameroon	5.151	1.159	–	1.033	4.884	2.076	2.827	3.828	1.091	4.932
Central African Rep.	2.422	1.568	1.033	–	3.334	1.449	1.282	1.574	0.153	2.127
Ethiopia	0.975	3.263	4.884	3.334	–	8.094	2.244	0.843	2.946	1.028
Iran	7.593	1.122	2.076	1.449	8.094	–	0.403	6.511	1.646	2.181
Liberia	1.049	0.834	2.827	1.282	2.244	0.403	–	2.073	1.094	2.178
Niger	0.848	3.434	3.828	1.574	0.843	6.511	2.073	–	1.537	0.527
Senegal	2.287	1.089	1.091	0.153	2.946	1.646	1.094	1.537	–	2.178
Somalia	0.101	3.540	4.932	2.127	1.028	2.181	2.178	0.527	2.178	–

Table C.2: *Matrix of D^2 values for a sample of ten countries*

with Ethiopia at $D^2 = 0.843$. As this is the lowest D^2 value for Ethiopia, a Ethiopia-Niger cluster is formed.

For these two pairs of education systems to link together, the Afghanistan-Somalia pair needs a value of 1.028 which is the maximum D^2 value among the four countries taken in pairs. As no other link involving these four countries together appears to be possible at a lower value, the two pairs join together at this D^2 value.

The emphasis on the maximum value for D^2 is the reason for this approach being called the 'diameter method'. It really involves determining the size for the diameter of the circle or sphere which encompasses two or more education systems. It is almost like asking how big a coin is needed to cover adjacent points on a graph. Hence in the above situation, a circle having a (squared) diameter[1] of 0.101 units is required to 'cover' both Afghanistan and Somalia. A circle of 0.527 units in (squared) diameter is needed to cover both Niger and Somalia but this still does not include Afghanistan to which Somalia is already joined. To do this, a circle (or sphere) with a (squared) diameter of 0.848 units is required.

A new starting point should now be found. A systematic way to work through the matrix is to select the first country not yet clustered. This is Burma. Its smallest D^2 value is with Liberia. The minimum value for Liberia, however, is with Iran ($D^2 = 0.403$) and this relationship is reciprocated. Hence, an Iran-Liberia pair is formed. Burma can now join this group at 1.122 but it possesses a value with Senegal which is lower than 1.122. This relationship is not reciprocal either, however, because the Senegal-Central African Republic link is only 0.153. As the Central African Republic has no lower value, this pair forms. For Burma to join both of these countries, the value for D^2 must now equal 1.568. A lower value for the pair exists, however, and the country which does join the Central African Republic-Senegal combination is Cameroon at $D^2 = 1.091$.

Examining the values for Burma again sets limits for D^2 as follows:

☐ 1.568 for the Cameroon group
☐ 1.122 for the Iran-Liberia combination
☐ 3.540 for the Afghanistan group.

As 1.122 is the next smallest value for the Iran-Liberia combination also, a triplet of countries is formed. (The smaller values for Liberia with Afghanistan and Senegal are not considered, of course, because of the combinations both of these other countries have previously entered into.)

To this stage, three groups have been formed. One of these contains four countries while the others each contain three countries. If clustering is to continue, the maximum D^2 values must be determined for each possible link. They are as follows:

---

[1] If D and not D^2 values were being used here, they would give the diameter measures directly.

- ☐ between the Afghanistan and Burma clusters, $D^2 = 8.094$
- ☐ between the Afghanistan and Central African Republic clusters, $D^2 = 5.151$
- ☐ between the Central African Republic and Burma clusters, $D^2 = 2.827$.

Hence the Central African Republic-Burma triplets join at $D^2 = 2.827$ and then all ten countries form into one group at $D^2 = 8.094$. Clustering loses its meaning if continued too far, however, and the very large jump in the D^2 value between the third last and the second last link points out that perhaps the clustering process should be stopped with the three clusters formed to a value of 1.122.

The results of this clustering can be presented as in Diagram 10.1. Although a scale of the magnitude of errors is used there for the horizontal axis, the diagram would look the same if D^2 values had been measured along the horizontal axis. The same clusters exactly would be formed. When computer facilities are not available, it is not really practical to use the slightly preferable 'error scale', and then D^2 (or D) values must be used.

Bibliography

Adams, D (1970) *Education and Modernization in Asia*, Reading, Mass: Addison-Wesley Publishing Co.

Adelman, I and Morris, C T (1967) *Society, Politics and Economic Development*, Baltimore: John Hopkins Press.

Adelman, I and Morris, C T (1972) The measurement of institutional characteristics of nations: methodological considerations, *Journal of Development Studies*, 8, 111-36.

Alker, H R (1965) *Mathematics and Politics*, New York: Macmillan.

Alker, H R and Russett, B M (1966) Indices for comparing inequality, *in* R L Merritt and S Rokkan, *Comparing Nations: the Use of Quantified Data in Cross-National Research*, New Haven: Yale University Press.

Almond, G A (1956) Comparative political systems, *Journal of Politics*, 18, 391-409.

Almond, G A and Verba, S (1963) *The Civic Culture: Political Attitudes and Democracy in Five Nations*, Princeton: Princeton University Press.

Anderson, C A (1961) Methodology of comparative education, *International Review of Education*, 7, 1-23.

Andrew, G M and Moir, R E (1970) *Information-Decison Systems in Education*, Illinois: F E Peacock Publishers Inc.

Angell, R C (1951) The computation of indexes of moral integration *in* P F Lazarsfeld and M Rosenberg (eds) *The Language of Social Research*, (1955), New York: The Free Press.

Arndt, H and Sundrum, S (1975) Regional price disparities, *Bulletin of Indonesian Economic Studies*, 11, 30-68.

Arriaga, E E (1970) A new approach to the measurement of urbanization, *Economic Development and Cultural Change*, 18, 206-18.

Asher, W and Shively, J E (1969) The technique of discriminant analysis: a reclassification of Harbison and Myers' seventy-five countries, *Comparative Education Review*, 13, 180-86.

Australian Bureau of Census and Statistics, *Schools* (various editions) Canberra: Australian Government Publications.

Babcock, L R and Nagda, N L (1972) Indices of air quality *in* W A Thomas (ed) *Indicators of Environmental Quality*, New York: Plenum Press.

Baker, B O *et al* (1966) Weak measurements vs strong statistics: an empirical critique of S S Stevens' proscriptions on statistics, *Educational and Psychological Measurement*, 26, 291-309.

Banghart, F W (1969) *Educational Systems Analysis*, London: Macmillan.

Banks, A S and Gregg, P M (1965) Grouping political systems: Q factor analysis of a 'Cross-Polity Survey', *The American Behavioural Scientist*, 9, 3-6.

Banks, A S and Textor, R B (1963) *A Cross-Policy Survey*, Cambridge, Mass: MIT Press.

Barrows, R L and Shaffer, R E (1975) Indicators of development in Wisconsin counties: 1970, *Social Indicator Research*, 2, 333-60.

Baster, N (1972) Development indicators: an introduction, *Journal of Development Studies*, 8, 1-20.

Beeby, C E (1966) *The Quality of Education in Developing Countries.* Cambridge, Mass: Harvard University Press.

Bereday, G Z F (1964) *Comparative Method in Education*, New York: Holt Rinehart and Winston Inc.

Bergson, A (1953) Reliability and usability of soviet statistics: a summary appraisal, *The American Statistician*, 19-23.

Berry, B J L (1960) An inductive approach to the regionalization of economic development, *in* N Ginsburg (ed) *Essays on Geography and Economic Development*, Dept. of Geography Research Paper No. 62, Chicago: University of Chicago Press.

Berstecher, D and Dieckmann, B (1969) On the role of comparisons in educational research, *Comparative Education Review*, 13, 96-103.

Biddle, B J (1970) Role conflicts perceived by teachers in four English-speaking countries, *Comparative Education Review*, 14, 30-44.

Biderman, A D (1971) Kinostatistics for social indicators, *Educational Broadcasting Review*, 5, 13-19.

Black, C E (1967) *The Dynamics of Modernization: A Study in Comparative History*, New York: Harper and Row.

Blalock, H M (1960) *Social Statistics*, New York: McGraw-Hill.

Blaug, M (1965) The rate of return on investment in education in Great Britain *The Manchester School*, 33, 205-51.

Blaug, M (1972) *An Introduction to the Economics of Education*, London: Penguin.

Bowles, S (1969) *Planning Educational Systems for Economic Growth*, Mass: Harvard University Press.

Bowman, M J (1951) A graphical analysis of personal income distribution in the United States, *American Economic Review*, 35, 607-28.

Bowman, M J and Anderson, C A (1963) Concerning the role of education in development *in* C Geertz (ed) *Old Societies and New States*, New York: The Free Press.

Brandis, W (1972) An index of Social Class, *in* W Brandis and D Henderson (1972) *Social Class, Language and Communication*, London: Routledge and Kegan Paul.

Bullock, A (1975) (ed) *A Language for Life*, London: HMSO.

Campbell, R W (1959) Problems of United States-Soviet economic comparisons, *in Comparisons of the United States and Soviet Economics*, Part I, Joint Economic Committee, Congress of the United States, 13-30.

Cant, R G (1975) Territorial socio-economic indicators in development plans in the Asian Region, *International Social Science Journal*, 27, 53-77.

Carr-Hill, R A (1973) A system of educational indicators with specific reference to population characteristics in developing countries. Paper presented to the Regional Seminar of Experts on Population Dynamics and Educational Planning, September 10-18, 1973, Bangkok, Unesco.

Carroll, J B and Chall, J S (1975) (eds) *Towards a Literate Society*, New York: McGraw-Hill.

Carter, L F (1947) An experiment in the design of tables and graphs for presenting numerical data, *Journal of Applied Psychology*, 21, 640-50.

Cattell, R B (1949) The dimensions of culture patterns by factorization of national characters, *Journal of Abnormal and Social Psychology*, 44, 443-69.

Cattell, R B (1950) The principal culture patterns discoverable in the syntal dimensions of existing nations, *The Journal of Social Psychology*, 32, 215-53.

Cattell, R B (1952) The three basic factor-analytic research designs – their inter-correlations and derivatives, *Psychological Bulletin*, 49, 499-520.

Cattell, R B (1978) *The scientific use of factor analysis in behavioural and life sciences*, Illinois: IPAT.

Cazes, B (1972) The development of social indicators: a survey, *in* A Shonfield and S Shaw (eds), *Social Indicators and Social Policy*, London: Heinemann Educational Books.

Chernoff, H (1973) The use of faces to represent points in K-dimensional space graphically, *Journal of the American Statistical Association*, 68, 361-68.

Chesswas, J D (1967) The basic data needed for educational planning *in* G Z F Bereday and J A Lauwerys (eds) *The World Yearbook of Education 1967: Educational Planning*, London: Evans Bros Ltd.

Chesswas, J D (1969) *Methodologies of Educational Planning for Developing Countries*, 2 vols, Paris, UNESCO, IIEP.

Choppin, B H (1974) The introduction of new science curricula in England and Wales, *Comparative Education Review*, 18, 196-206.

Coale, A J and Stephan, F F (1962) The case of the Indians and the teenage widows, *Journal of the American Statistical Association*, 57, 338-47.

Cohen, W J (1967) Education and learning, *Annals of the American Academy of Political and Social Science*, 373, Sept. 79-101.

Cohen, W J (1975) Educational indicators and social policy *in* ETS *Educational Indicators: Monitoring the State of Education*, Proceedings of the 1975 ETS Invitational Conference, Princeton, NJ: Educational Testing Service.

Coleman, J S (1964) *Introduction to Mathematical Sociology*, Glencoe: The Free Press.

Coleman, J S (1969) The methods of sociology *in* R Bierstedt *A Design for Sociology: Scope, Objectives, Methods*, Philadelphia: The American Academy of Political and Social Sciences.

Colley, D G (1975) Social change index – an objective means to discern and measure the relative current social condition of cities, towns and their sub-communities, *Social Indicators Research*, 2, 93-118.

Collins, W A (1970) Interviewers' verbal idiosyncrasies as a source of bias, *Public Opinion Quarterly*, 34, 416-22.

Conrad, A H (1954) Redistribution through government budgets in the United States, 1950 *in* A T Peacock (ed) *Income Redistribution and Social Policy*, London: Jonathan Cape, 178-267.

Consolazio, W V (1967) The Dynamics of Academic Science: A Degree Profile for Academic Science and Technology and the Contributions of Federal Funds to Universities and Colleges, Research Paper NSF-67-6, Washington: National Science Foundation.

Cooley, W W and Lohnes, P R (1971) *Multivariate Data Analysis*, New York: John Wiley.

Coombs, P H (1968) *The World Educational Crisis: a Systems Analysis*, Oxford University Press, New York.

Coombs, P H (1969) Time for a change of strategy *in* C E Beeby (ed) *Qualitative Aspects of Educational Planning*, Paris: Unesco, IIEP.

Correa, H (1975) Quantitative analysis of the implementation of educational plans in Latin America, *Socio-Economic Planning Sciences*, 9, 247-55.

Crowe, W R (1965) *Index Numbers: Theory and Applications*, London: MacDonald & Evans Ltd.

Croxton, F E and Stein, H (1932) Graphic comparisons by bars, squares, circles and cubes, *Journal of the American Statistical Association*, 27, 54-60.

Culyer, A J *et al* (1972) Health indicators *in* A Shonfield and S Shaw (eds) *Social Indicators and Social Policy*, London: Heinemann Educational Books.

Cutright, P (1963) National political development: measurement and analysis, *American Sociological Review*, 28, 253-64.

Dahl, R A (1967) The evaluation of political systems *in* C L Taylor, *Aggregate Data Analysis: Political and Social Indicators in Cross-National Research*, Paris: Mouton and Co, 47-57.

Dalkey, N C and Rourke, D L (1971) *Experimental Assessment of Delphi Procedures with Group Value Judgments*, Rand Report R-612, Santa Monica, California: Rand Corporation.

Darlington, R B *et al* (1973) Canonical variate analysis and related techniques, *Review of Educational Research*, 43, 433-54.

Davis, J A (1964) *Great aspirations: The Graduate School Plans of America's College Seniors*, Chicago: Aldine Publishing Co.

Deming, W E (1950) *Some Theory of Sampling*, New York: John Wiley.

Denison, E F (1964) Measuring the contribution of education (and the residual) to economic growth *in The Residual Factor and Economic Growth*, Paris: OECD, 13-85.

Dewey, C (1974) The agricultural statistics of the Punjab 1867-1947, *Bulletin of Quantitative and Computer Methods in South Asian Studies*, 3, 3-14.

Drewnowski, J (1970) *Studies in the Measurement of Levels of Living and Welfare*, Report No. 70. 3, Geneva, UNRISD.

Drewnowski, J (1972) Social indicators and welfare measurements – remarks on methodology, *Journal of Development Studies*, 8, 77-90.

Drewnowski, J and Subramanian, M (1970) Social aims in development plans, *in Studies in the Methodology of Social Planning*, Report No. 70. 5, Geneva, UNRISD.

Dubin, R (1969) *Theory Building*, New York: The Free Press.

Duncan, B (1968) Trends in output and distribution of schooling *in* E B Sheldon and W E Moore (eds) *Indicators of Social Change: Concepts and Measurement*, New York: Russell Sage Foundation, 601-72.

Dyer, A S *et al* (1969) A comparison of four methods of obtaining discrepancy measures based on observed and predicted school system means on achievement tests, *American Educational Research Journal*, 6, 591-605.

Ebel, R L (1969) The relation of scale fineness to grade accuracy, *Journal of Educational Measurement*, 6, 217-21.

Ebel, R L (1972) *Essentials of Educational Measurement*, Englewood Cliffs, NJ: Prentice-Hall.

Edding, F (1962) The economics of educational planning, *International Review of Education*, 8, 93-97.

Edwards, R (1970) Presidential address: The dimensions of comparison, and of comparative education, *Comparative Education Review*, 14, 239-54.

Eide, K (1973) Approaches to educational planning in a population and societal context. Paper delivered to the Regional Seminar of Experts on Population Dynamics and Educational Planning, September 10-18, Bangkok, Unesco BKM/73/D/347-80.

Etzioni, A and Lehman, E W (1969) Some dangers in valid social measurement, *in* B M Gross, *Social Intelligence for America's Future: Explorations in Societal Problems*, Boston: Allyn and Bacon.

Fanchette, S (1974) Social indicators: problems of methodology and selection, *in Social Indicators: Problems of Definition and of Selection*, Paris, Unesco.

Farrell, J P (1969) The structural differentiation of developing educational systems: a Latin American comparison, *Comparative Education Review*, 13, 294-311.

Feinberg, B M (1972) *Kinostatistics: Communicating a Social Report to the Nation*, Washington: Bureau of Social Science Research Inc, Report No. 532.

Fennessey, A J (1968) The general linear model: New perspectives, *American Journal of Sociology*, 74, 1-27.

Ferriss, A L (1969) *Indicators of Trends in American Education*, New York: Russell Sage Foundation.

Finlayson, D S *et al* (1971) *Administration Manual for the School Climate Index*, Slough: National Foundation for Educational Research.

Fowler, R F (1970) *Some Problems of Index Number Construction*, Studies in Official Statistics, Research Series No. 3, London: HMSO.

Fox, T G and Miller, S M (1965) Economic, political and social determinants of mobility: An international cross-sectional analysis, *Acta Sociologica*, 9, 76-93.

Fuller, W P (1971) A review of statistics for planning education in Thailand, Bangkok: National Education Council.

Gallup, G H (1969-1976) Annual Gallup Poll of the Public's Attitudes toward the Public Schools, *Phi Delta Kappan*, 51-58.

Garms, W I Jr (1968) The correlates of educational effort: A multivariate analysis, *Comparative Education Review*, 12, 281-99.

Gibb, C A (1956) Changes in the culture pattern of Australia, 1906-1946, as determined by p-technique, *Journal of Social Psychology*, 43, 225-38.

Glock, C Y and Stark, R (1965) *Religion and Society in Tension*, Chicago: Rand-McNally.

Gordon, R (1968) Issues in multiple regression, *American Journal of Sociology*, 73, 592-616.

Greenberg, J H (1956) The measurement of linguistic diversity, *Language*, 32, 109-15.

Gregg, P M and Banks, A S (1965) Dimensions of political systems: factor and analysis of 'A Cross-Polity Survey', *American Political Science Review*, 59, 602-14.

Gully, W H and Newton, C H (1972) Methods of measuring the distribution of socio-economic conditions, *Socio-Economic Planning Sciences*, 6, 187-96.

Hall, M and Tideman, N (1967) Measures of Concentration, *Journal of the American Statistical Association*, 62, 162-68.

Halpin, A W and Croft, D B (1965) *The Organizational Climate of Schools*, Chicago: Midwest Admin. Centre, University of Chicago.

Harbison, F and Myers, C A (1964) *Education, Manpower and Economic Growth: Strategies of human resource development*, New York: McGraw-Hill.

Harbison, F *et al* (1970) *Quantitative Analyses of Modernisation and Development*, Research Report No. 15, Princeton: NJ: Industrial Relations Section, Department of Economics, Princeton University.

Harper, J and Richer, S (1969) Sociology of Education, *Review of Educational Research*, 39, 671-94.

Hobson, R and Mann, S H (1975) A social indicator based on time allocation, *Social Indicators Research*, 1, 439-57.

Hopper, E I (1968) A typology for the classification of educational systems, *Sociology*, 2, 29-46.

Horn, R V (1975) Social indicators for development planning and analysis, *International Labour Review*, 111, 483-506.

Huff, D (1954) *How to Lie with Statistics*, London: Oliver and Boyd.

Husek, T R and Sirotnik, K (1968) Matrix sampling, *Evaluation Comment*, 1, 1-4.

Husén, T (ed) (1967) *International Study of Achievement in Mathematics*, Vols I and II, Stockholm: Almquist and Wiksell, New York: John Wiley.

Ivanovic, B (1974) A method of establishing a list of development indicators, *in Social Indicators: Problems of definition and selection*, Paris: Unesco.

Jahn, J *et al* (1947) The measurement of ecological segregation, *American Sociological Review*, 12, 293-303.

Jahoda, G (1961) Aspects of Westernization: A study of adult-class students in Ghana: I, *British Journal of Sociology*, 12, 375-86.

Jencks, C *et al* (1972) *Inequality: A Reassessment of the Effectiveness of Family and Schooling in America*, New York: Basic Books.

Johnson, S C (1967) Hierarchical clustering schemes, *Psychometrika*, 32, 241-54.

Johnstone, J N (1974) Mathematical models developed for use in educational planning: a review, *Review of Educational Research*, 44, 177-201.

Johnstone, J N (1976) Typology formation across socio-economic indicators, *Socio-Economic Planning Sciences*, 10, 167-71.

Johnstone, J N (1977) The dimensions of educational systems, *Comparative Education Review*, 21, 51-68.

Johnstone, J N (1978a) *Indicators of Indonesian Provincial Education Systems*, Jakarta: Ministry of Education and Culture.

Johnstone, J N (1978b) Education systems: approaches and methods in their evaluation, *Evaluation in Education: International Progress*, Vol. 2, Monograph 3.

Johnstone, J N and Tobing, P (1976) *The Indonesian Educational Simulation Model*, Jakarta: Ministry of Education and Culture.

Kamrany, N W and Christakis, A N (1970) Social indicators in perspective, *Socio-Economic Planning Sciences*, 4, 207-16.

Kartomo, W and Heckelman, T J (1971) The systems approach to educational reform in Indonesia, *Bulletin of the Unesco Regional Office for Education in Asia*, 6, 25-39.

Keeves, J P *et al* (1976) *Australian Studies in School Performance*, 3 Vols., ERDC Reports 8, 9 (1976) and 10 (1977), Canberra: Australian Government Publishing Service.

Kendall, M G and Stuart, A (1958) *The Advanced Theory of Statistics*, Vol. 1, London: Charles Griffin and Co.

Kerlinger, F N and Pedhazur, E J (1973) *Multiple Regression in Behavioural Research*, New York: Holt, Rinehart and Winston Inc.

King, M A (1972) Primary and secondary indicators of education, *in* A Shonfield and S Shaw *Social Indicators and Social Policy*, London: Heinemann Educational Books.

Kingdom of Saudi Arabia (1970) *Development Plan 1390 A.H.*, Kingdom of Saudi Arabia, Central Planning Organisation.

Kottis, A P and Kottis, G C (1973) Educational disparities in an egalitarian society: the case of Norway, *International Review of Education*, 19, 385-95.

Land, K C (1975) Theories, models and indicators of social change, *International Social Science Journal*, 27, 7-37.

Lave, R E Jr and Kyle, D W (1968) The application of systems analysis to educational planning, *Comparative Education Review*, 12, 39-55.

Lazarsfeld, P F (1959) Problems in methodology *in* R K Merton *et al Sociology Today*, New York: Basic Books.

Lazarsfeld, P F and Barton, A H (1951) Qualitative Measurement in the Social Sciences: Classification, typologies, and indices, *in* D Lerner and H D Lasswell (eds) *The Policy Sciences*, Stanford: Stanford University Press.

Levinson, L L (1963) *The Left-Handed Dictionary*, New York: Collier Books.

Lindgren-Hooker, B (1974) *Educational Flow Models: Applications to Arab Statistical Data*, Linkoping Studies in Education, Dissertations No. 3, Linkoping, Sweden.

Little, A and Mabey, C (1972) An index for designation of Educational Priority Areas *in* A Shonfield and S Shaw, *Social Indicators and Social Policy*, London: Heinemann Educational Books, 67-93.

Lord, F M (1953) On the statistical treatment of football numbers, *American Psychologist*, 8, 750-51.

McClelland, D C (1966) Does education accelerate economic growth? *Economic Development and Cultural Change*, 14, 257-78.

McGranahan, D V *et al* (1972) *Contents and Measurement of Socio-economic Development*, New York: Praeger.

McNamara, J F (1971) Mathematical programming models in educational planning, *Review of Educational Research*, 41, 419-46.

McQuitty, L L (1966) Single and multiple hierarchical classification by reciprocal pairs and rank order types, *Educational and Psychological Measurement*, 26, 253-65.

McQuitty, L L (1971) A comparative study of some selected methods of pattern analysis, *Educational and Psychological Measurement*, 31, 607-26.

Marco, G L (1974) A comparison of selected school effectiveness measures based on longitudinal data, *Journal of Educational Measurement*, 11, 225-34.

Mehrens, W A and Lehmann, I J (1973) *Measurement and Evaluation in Education and Psychology*, New York: Holt, Rinehart and Winston.

Meyer, J W (1971) Economic and political effects on national educational enrolment patterns, *Comparative Education Review*, 15, 28-43.

Monkhouse, F J and Wilkinson, H R (1963) *Maps and Diagrams*, London: Methuen and Co.

Mushkin, S J and Billings, B B (1975) Measures of educational outcomes in developing countries *in Educational Indicators: Monitoring the State of Education*, ETS: Princeton, New Jersey.

Noah, H J and Eckstein, M A (1969) *Toward a Science of Comparative Education*, Toronto: The Macmillan Co.

OECD (1967) *Methods and Statistical Needs for Educational Planning*, Paris: OECD.

OECD (1971) Indicators of Educational Performance, Report DAS/EID/71.13, Paris: OECD.

OECD (1973a) *A Framework for Educational Indicators to Guide Government Decisions*, Paris: OECD.

OECD (1973b) *Indicators of Performance of Educational System*, Paris: OECD.

OECD (various years) *Classification of Educational Systems in OECD Countries*, Paris: OECD.

Orleans, L A (1974) Chinese statistics: the impossible dream, *The American Statistician*, 28, 47-52.

Parsons, R (1974) *Statistical Analysis: a Decision Making Approach*, New York: Harper and Row.

Passow, A H *et al* (1974) *The National Case Study: An Empirical Comparative Study of Twenty-One Educational Systems*, Stockholm: Almquist and Wiksell.

Peaslee, A L (1969) Education's role in development, *Economic Development and Cultural Change*, 17, 293-318.

Philp, H (1969) The evaluation of quality in education, *in* C E Beeby (ed), *Qualitative Aspects of Educational Planning*, Paris: Unesco, IIEP, 280-91.

Pinchak, B M (1974) Regional Differences in Grading Practices, ETS Research Bulletin RB-74-22, Princeton, NJ: Educational Testing Service.

Poignant, R (1972) Possible criteria for evaluating education policies, *in Evaluation of Educational Systems*, Paris: Unesco/IIEP.

Postlethwaite, T N (1974) Target populations, sampling, instrument construction and analysis procedures, *Comparative Education Review*, 18, 164-79.

Psacharopoulos, G (1972) Rates of return to investment in education around the world, *Comparative Education Review*, 16, 54-67.

Psacharopoulos, G (1975) *Earnings and Education in OECD countries*. Paris: OECD.

Rae, D W and Taylor, M (1970) *The Analysis of Political Cleavages*, New Haven: Yale University Press.

Retzlaff, R H (1965) The use of aggregate data in comparative political analysis *in* C L Lewis (1968) *Aggregate Data Analysis: Political and Social Indicators in Cross-National Research*, Paris: Morton and Co, 63-78.

Robinson, W S (1950) Ecological correlations and the behaviour of individuals, *American Sociological Review*, 15, 351-57.

Rummel, R J (1970a) Dimensions of error in cross-national data *in* R Naroll and R Cohen (eds) *A Handbook of Method in Cultural Anthropology*, New York: The National History Press.

Rummel, R J (1970b) *Applied Factor Analysis*, Evanston: Northwestern University Press.

Rummel, R J (1972) *The Dimensions of Nations*, Beverly Hills, California: Sage Publications.

Russett, B M (1965) A note on the evaluation of error and transformation in data analysis, *The American Political Science Review*, 59, 444-46.

Russett, B M (1968) Delineating international regions *in* J D Singer (ed) *Quantitative International Politics: Insights and Evidence*, New York: The Free Press.

Russett, B M *et al* (1964) *World Handbook of Political and Social Indicators*, New Haven: Yale University Press.

Scheuch, E K (1966) Cross-national comparisons using aggregate data: Some substantive and methodological problems, *in* R L Merritt and S Rokkan (eds) *Comparing Nations: The Use of Quantitative Data in Cross-National Research*, New Haven and London: Yale University Press.

Schnore, L F (1961) The statistical measurement of urbanization and economic development, *Land Economics*, 37, 230-45.

Schuman, H and Converse, J M (1971) The effects of black and white interviewers on black responses in 1968, *Public Opinion Quarterly*, 35, 44-68.

Schwarz, W (1969) Liberia *in* C Legum (ed), *Africa Handbook*, Middlesex: Penguin Books Ltd.

Seers, D (1972) What are we trying to measure?, *Journal of Development Studies*, 8, 21-36.

Sen, A (1973) *On Economic Inequality*, New York: W W Norton & Co.

Shapiro, M J (1970) Discovering interviewer bias in open-ended survey responses, *Public Opinion Quarterly*, 34, 412-15.

Shoemaker, D M (1975) Toward a framework for achievement testing, *Review of Educational Research*, 45, 127-47.

Shults, W D and Beauchamp, J J (1972) Statistically based air-quality indices *in* W A Thomas (ed) *Indicators of Environmental Quality*, New York: Plenum Press.

Sokal, R R and Sneath, P H A (1963) *Principles of Numerical Taxonomy*, San Francisco, Cal: Freeman.

Spurr, W A and Bonini, C P (1967) *Statistics for Business*, Holmwood, Ill: Irwin.

Stake, R E (1970) National Assessment, *in* Proceedings of the 1970 Invitational Conference on Testing Problems, *The promise and perils of educational information systems*, Princeton, NJ: Educational Testing Service.

Stern, L N *et al* (1973) On the dimensions of political culture: a new perspective, *Comparative Political Studies*, 5, 493-512.

Stone, R (1971) *Demographic Accounting and Model-Building*, Paris: OECD.

Sudrapraesert, K and Tunscri, V (1977) Regional disparities in the development of education in Thailand, Paper presented at IIEP Conference on Regional Disparities, December, 1977.

Svennilson, I *et al* (1962) *Targets for Education in Europe in 1970*, Policy Conference on Economic Growth and Investment in Education, Paris: OECD.

Tatsouka, M M and Tiedman, D V (1963) Statistics as an aspect of scientific method in research on teaching *in* N L Gage (ed) *Handbook of Research in Teaching*, Chicago: Rand McNally and Co.

Taylor, C L and Hudson, M C (1972) *World Handbook of Political and Social Indicators*, 2nd edn, New Haven: Yale University Press.

Terwilliger, J S (1971) *Assigning Grades to Students*, Glenview: Scott, Foresman and Co.

Terwilliger, J S and Anderson, D H (1969) An Empirical Study of the Effects of Standardizing Scores in the Formation of Linear Composites, *Journal of Educational Measurement*, 6, 145-54.

Thias, H H and Carnoy, M (1969) *Cost-Benefit Analysis in Education: A Case Study on Kenya*, IBRD Report EC-173, Washington: IBRD.

Thonstad, T (1969) *Education and Manpower: Theoretical Models and Empirical Applications*, London: Oliver and Boyd.

Turnbull, W W (1975) Preface to the Conference, *Educational Indicators: Monitoring the State of Education*, Proceedings of the 1975 ETS Invitational Conference, Princeton: ETS.

Turner, R H (1960) Contest and Sponsored Mobility and the School System, *American Sociological Review*, 25, 855-67.

Unesco (1955) *World Survey of Education I: Handbook of Educational Organization and Statistics*, Paris: Unesco.

Unesco (1966) *An Asian Model of Educational Development: Perspectives for 1965-80*, Paris: Unesco.

Unesco (1974) A system of educational indicators, Paris: Unesco (Mimeo) ED.75/WS.8, December 1974.

Unesco (1976) *International standard classification of education*, COM/ST/ISCED, Paris: Unesco.

Unesco *Statistical Yearbook* (various editions) Paris: Unesco.

United Nations *Statistical Yearbook* (various editions), New York: UN.

United States Department of Health, Education and Welfare (1969) *Toward a Social Report*, Washington, DC: US Government Printing Office.

Wanderer, J J (1969) An index of riot severity and some correlates, *American Journal of Sociology*, 74, 500-05.

Warren, J R (1971) *College Grading Practices: an Overview*, Report No. 9 Washington, DC: ERIC Clearinghouse on Higher Education.

Welsh, W A (1971) Toward effective typology construction in the study of Latin American political leadership, *Comparative Politics*, 3, 271-80.

Williamson, J G (1965) Regional inequality and the process of national development: a description of the patterns, *Economic Development and Cultural Change*, 13, 3-45.

World Bank (1979) *Atlas 1979*, Washington: World Bank.

Yeh, S H K (1974) Some observations on the use of social indicators in development planning, with special reference to levels of living. Paper presented to the Working Group on Applicability of Socio-Economic Indicators to Development Planning, Bangkok, September 16-20, 1974, Unesco.

Names Index

Adams, D, 25
Adelman, I, 31, 38, 70, 130, 285
Alker, H R, 92, 95
Almond, G A, 11, 257
Anderson, C A, 11, 257, 285
Anderson, D H, 80
Andrew, G M, 25
Angell, R C, 77, 87, 170
Arndt, H, 210
Arriaga, E E, 99, 100
Asher, W, 142

Babcock, L R, 117
Baker, B O, 48
Banghart, F W, 25
Banks, A S, 37, 49, 52, 130, 246
Barrows, R L, 196-8, 202-3, 209
Barton, 43
Baster, N, 9, 22
Beauchamp, J J, 77
Beeby, C E, 257, 258-9
Bereday, G Z F, 257
Bergson, A, 57
Berry, B J L, 246, 257
Berstecher, D, 257
Biddle, B J, 148
Biederman, A D, 206
Black, C E, 7
Blalock, H M, 129
Blaug, M, 138, 179
Bonnini, C P, 131
Bowles, S, 95
Bowman, M J, 11, 36, 94
Brandis, W, 69, 130
Bullock, A, 168

Campbell, R W, 57
Cant, R G, 37
Carnay, M, 179
Carr-Hill, R A 137
Carroll, J B, 168
Carter, L F, 206
Cattell, R B, 70, 125, 246, 260
Cazes, B, 9
Chall, J S, 168
Chernoff, H, 220
Chesswas, J D, 21
Choppin, B H, 168
Christakis, A N, 31, 70
Coale, A J, 67
Cohen, W J, 4, 8
Coleman, J S, 6, 67, 73
Colley, D G, 9
Collins, W A, 66
Conrad, A H, 280
Consolazio, W V, 77
Converse, J M, 62
Cooley, W W, 247
Coombs, P H, 9, 25, 54, 285
Correa, H, 116-17
Croft, D B, 189
Crowe, W R, 75, 106
Croxton, F E, 209
Culyer, A J, 9, 29
Cutright, P, 11, 37, 84, 88

Dahl, R A, 23
Dalkey, N C, 76
Darlington, R B, 247
Davis, J A, 170
Deming, W E, 63
Denison, E F, 179
Dewey, C, 61, 62, 63, 68
Dieckmann, B, 257
Drewnowski, J, 7, 25, 29, 31
Dubin, R, 6
Duncan, B, 202
Dyer, H S, 201

Ebel, R L, 49, 82
Eckstein, M A, 250, 253
Edding, F, 11, 249

Edwards, R, 15
Eide, K, 54
Etzioni, A, 55-6, 58, 69

Fanchette, S, 31, 137
Farrell, J P, 42, 134, 135, 162
Feinberg, B M, 205-6
Fennessey, A J, 41
Ferriss, A L, 8, 17, 31, 42, 76, 96, 168-70
Finlayson, 189
Fowler, R F, 106
Fox, T G, 11
Fuller, W P, 137

Gallup, G H, 148, 172-4
Gibb, C A, 12
Glock, C A, 12, 47
Gordon, R, 41
Greenberg, J H, 100
Gregg, P M, 37, 130, 246
Gulley, W H, 76

Hall, M, 97, 99
Halpin, A W, 189
Harbison, F, 11, 21, 31, 37, 76, 86, 116, 119, 138, 249, 259
Harper, J, 134, 157
Heckelman, T J, 25
Hobson, R, 76
Hopper, E I, 258-9
Horn, R V, 22
Hudson, M C, 52, 95, 97, 100
Huff, D, 75
Husek, T R, 200
Husen, T, 168

Ivanovic, B, 100

Jahn, J, 92, 95
Jahoda, G, 11
Jencks, C, 91
Johnson, S C, 261
Johnstone, J N, 21, 25, 63, 84, 110, 138, 149, 151, 234, 247, 260, 263, 269, 277, 285

Kamrany, N W, 31, 70
Kartomo, W, 25
Keeves, J P, 168
Kendall, M G, 280
Kerlinger, F N, 132, 247
King, M A, 157
Kottis, A P & G C, 96
Kyle, D W, 25

Land, K C, 2
Lave, R E jr, 25
Lazarsfeld, P F, 22, 34, 43, 58
Lehmann, 55-6, 58, 69, 82
Levinson, L L, 81
Lindgren-Hooker, B, 157
Little, A, 84, 195-6
Lohnes, P R, 247
Lord, F M, 48, 78

Mabey, C, 84, 195-6
Mann, S H, 76
Marco, G L, 201-2
McClelland, D C, 11, 249
McGranahan, D V, 31, 38, 54, 70, 130, 139, 216-17, 285
McNamara, J F, 277
McQuitty, L L, 260
Mehrens, W A, 82
Meyer, J W, 11
Miller, S M, 11
Moir, R E, 25
Monkhouse, F J, 207
Morris, C T, 21, 31, 38, 70, 130, 285
Mushkin, S J, 30
Myers, C A, 11, 21, 37, 76, 81, 119, 138, 249, 259

Nagda, N L, 117
Newton, C H, 76
Noah, H J, 250, 253

OECD, 21, 25, 37, 52, 137, 175
Orleans, L A, 57

Parsons, R, 106
Passow, A H, 168
Peaslee, A L, 31, 285, 295
Pedhazur, 132, 247
Philp, H, 55
Pinchak, B M, 170
Poignant, R, 28
Postlethwaite, T N, 168
Psacharopoulos, G, 179-80

Rae, D W, 100
Retzlaff, R H, 17, 68
Richer, S, 134, 157
Robinson, W S, 40
Rourke, D L, 76
Rummel, R J, 11, 12, 67, 70, 84, 125, 130, 246, 260
Russett, B M, 52, 67, 92, 95, 130, 246-55

Scheuch, E K, 41, 49
Schnore, L F, 15
Schuman, H, 62
Schwarz, W, 67
Seers, D, 8
Shaffer, R E, 196-8, 202-3, 209
Shapiro, M J, 66
Shively, J E, 142
Shoemaker, D N, 200
Shults, W D, 77
Sirotnik, K J, 200
Sneath, P H A, 260
Sokal, R R, 260
Spurr, W A, 131
Stake, R E, 31
Stark, R, 12, 47
Stein, H, 209
Stephan, F F, 67
Stern, L N, 257
Stone, R, 21, 25
Subramanian, M, 29, 36
Sudrapraesert, K, 90
Sundrum, S, 210
Svennilson, I, 11

Tatsouka, M M, 12
Taylor, C L, 52, 95, 97, 100
Taylor, M, 100
Terwilliger, J S, 80, 82
Textor, R B, 49, 52
Thias, H H, 179
Thonstad, T, 110
Tideman, N, 97, 99
Tiedman, D V, 12
Tobing, P, 110
Tunscri, V, 90
Turnbull, W W, 10
Turner, R H, 259

Unesco, 25, 52, 143-7, 151, 156-62, 176-8, 187, 258

Verba, S, 11, 257

Wanderer, J J, 134
Warren, J R, 170
Welsh, W A, 257
Wilkinson, 207
Williamson, J G, 91
World Bank, 213

Yeh, S H K, 31

Subject Index

Absolute measurement, 20, 211
Academic production index, 169-70
Achievement levels, 32-3, 57, 96, 117, 167-8, 191, 202
Addition, 47, 73-5
Africa, 46, 79-81, 162-3
Aggregate data, 42-3, 190
Agriculture, 68, 251, 256
Air quality index, 77, 117
Analysis of variance, 48
Arab countries, 157
Argentina, 19, 117, 119, 249
Asia, 46, 97, 134-5, 162-3
Australia, 52, 198, 220
Austria, 220
Availability of student places, 151-2, 226-8, 235-7, 276

Bar chart, 207-9
Base year, 102, 226
Bhutan, 20
Bolivia, 19, 117
Brazil, 19, 220
Budget, % for education, 11, 28, 142, 229
Buildings, 42, 47, 153
Bulgaria, 102
Burma, 265-7

C index of concentration, 20, 97-102, 114, 154-6, 158, 163, 171, 198, 230-1, 237-8
Canonical correlation, 131, 247
Census, 50-1
Centralization, 14, 42, 163
Ceylon, 267
Change, measurement of, 101, 181-6
Chile, 19, 267
China, 57
Class size, 118
Classes, relative availability of, 152, 235-7, 240
Classification of systems, *see* Typologies
Classroom interaction, 12, 189
Cluster analysis, 37, 138-41, 258-62
Coefficient of Variation, 90-1
Collective units, 40-2
Columbia, 19, 79-80, 220, 252
Communist countries, 249-51, 256
Community involvement, 85, 132, 189
Composite indices, 15, 16-17, 31, 70, 137, 275
Computer cards, 67
Concentration, measurement of, *see* C index
Concept substitution, 57
Consumer Price Index, 54, 166
Continuation coefficient, 109-10
Correlation coefficient, 48, 73, 78, 121, 192, 240-1, 260
Correspondence points, 216-18
Costs, *see* Education Cost Index
Crime Index, 78, 170
Criterion-referencing, 18, 23, 116-18, 276
Cross-Polity Survey, 49
Cubes, 207-9
Curriculum index, 196
Cyprus, 97-8, 155, 208
Czechoslovakia, 119

Data:
  comparability of, 51, 52, 56, 58, 187-8
  general, 42-5
  sources of, 50, 51-2
Data collection:
  annual, 62, 63
  errors in, 62-9
  methods, 63-6
  purposes, 39, 243

Delphi technique, 76
Denmark, 119, 248
Diachronic, *see* Self-referencing
Disaggregated index, 15-17
Distance, measures of, 116, 260-2, 297-301
Distribution of quantities, 21-2, 42-4, 88-101, 115, 158, 230-4, 238-9
Distributive indicators – defined, 21, 43
Division, *see* Multiplication

Ecological fallacy, 40-2
Ecuador, 19, 79-80
Educational pyramid, 151, 227
Education Cost Index, 104-7, 166
Education Product Index, 106-7, 166
Education system – definition, 24-33
Efficiency, 161, 232-4, 238
Egypt, 79-80, 208
El Salvador, 92-5, 282
Empirical definition (of indicators), 34, 36-8, 149-55
England, 157, 250
Enrolment ratio, 11, 15, 18, 20, 28, 60, 69, 90, 126, 138-9, 226
Entrants who graduate, Index of, 113-14, 171
Environmental quality – measurement of, 117
Equality – measurement of, 90-6, 159, 238-9
Errors in data, 49, 59-69
Ethiopia, 220
Evaluation, 167, 174, 191, 242
Expert judgement, 34, 36, 58, 76

Faces, 220-2
Factor analysis, 36, 125-31, 140-4, 193, 196, 202, 216, 285-96
Females, Involvement of, 17, 79-80, 95, 153-4, 208, 227-8, 237, 251-3
Financial Involvement Index, 142-4, 228-9
Finland, 220
Flow statistics, 21
Ford Foundation, 13, 51
Fractionalization, Index of, 99
Fractional measurement, 51, 55-6
France, 103
Funds – source of, 45, 163, 166

Gallup Polls, 107-8, 172-4
General education, 154, 220, 230-1, 237, 249
Germany, Federal Republic, 67, 119, 252
Gini coefficient, 20, 94-6, 115, 161, 176, 239, 280-4
Global data, 44
GNP (% on education), 11, 15, 32, 43, 229
GNP per capita, 209-11
Goals, *see* Objectives
Grade point average, 170
Grades, expected number, 109, 112-13, 117, 157-8, 233, 239
Graduation rates, 57, 171, 191
Graphs, 206-7
Guatemala, 268
Guttman Scale Analysis, 37, 133-4, 162
Guyana, 19, 79-80

Harbison and Myers' Index, 81, 119-21, 138-9
Health sector, 29, 31, 38, 175
Herfindahl-Hirschman index, *see* C index
Honduras, 220, 249
Human Resource Involvement Index, 102, 138-40, 225-6, 275

Ideal country, 116
IEA, 167-8, 199
Illiteracy, *see* Literacy (levels of)
Implementation, *see* Plan Implementation
Income, 66, 91, 95
Index numbers, 102-8, 152, 169, 178, 225-6, 235
India, 13, 20, 68, 119, 253
Indicators:
    characteristics, 2-6
    definition of, 2, 137
Indirect measurement, 50-2, 198
Individual units, 40-2
Indonesia, 22, 134, 192, 209-10, 234-42, 269-72
Inequality, *see* Equality
Inputs:
    definition of, 25-8
    indicators of, 69, 87, 124, 138-50, 163, 225, 247 ff, 262-8
Interchangeability, 58, 70, 118-21, 131, 138, 162, 202, 240-2
Interval scales, 46-50, 78, 118, 122, 133
Interviewer, 66
Iowa Basic Skills Tests, 198
IQ, 16, 31
Iran, 220, 268

Iraq, 119, 268
Israel, 220, 250
Italy, 67, 249, 268

Jamaica, 119
Japan, 97-8, 208, 250, 252

Karachi Plan, 13, 246
Kenya, 104-6
Kinostatistics, 245
Korea, Republic of, 250, 252, 253

Labour force, 175
Laos, 97-9, 134, 155
Laspeyres formula, 105-7, 149, 174
Lebanon, 119
Leisure time, index of, 76
Level of living index, 31
Likert scales, 133, 172
Line of equality, *see* Lorenz curve
Literacy:
  levels of, 40, 49, 61, 172
  rates, 11, 15, 85
Lorenz curve, 92-6
  see also Gini coefficient

Maldive Islands, 20
Mass media, 178, 181
Mathematical models, 68, 106-7, 277
Matrix algebra, 110-13
Matrix sampling, 200
Mean, 43, 89-90
  arithmetic, 93
  geometric, 93, 152
Measurement:
  scales, 45-50, 57
  units of, 40-2
  see also Interval, Nominal, Ordinal and Ratio scales
Median, 43
Mexico, 119
Missing data, 83
Mode, 43
Moderation, 132-3
Monitoring, 8-9, 101, 174, 234-40
Multiple discriminant analysis, 130
Multiplication, 47, 73-5

NAEP, 168, 199-200
National education systems, 30, 40, 42
Need achievement, 57
Netherlands, the, 252
New Zealand, 37
Nicaragua, 92-5, 114-15, 268
Nominal scales, 45-6, 47-50, 55, 78, 96, 99, 118, 122, 132
Non-distributive indicators:
  definition of, 22, 43
  development of, 78-85
Non-parametric statistics, *see* Parametric statistics
Normal curve, 82
Normative comment of indicators, 14, 188, 219, 275
Norm-referencing, 17-18, 20, 23, 78-101, 102, 123 ff, 207, 242-3, 276
Norway, 96

Objectives:
  indicators of, 106, 167-8
  statement of, 7, 24, 224-5, 243, 275
OCDQ, 189
OECD, 37, 51, 137, 175
Operational definition, 55-8
Ordinal scales, 44, 46, 47-50, 78, 96, 118, 122, 132
Orientation of studies, 154, 230-3, 237-8
Outputs:
  definition of, 25-7
  indicators of, 57, 87, 167-75, 198-203

Paasche formula, 104-6
Panama, 249, 252
Paraguay, 19, 252, 267
Parametric statistics, 48
Perception of central educational problems index, 108
Peru, 19, 119
Pie-chart, 207-8
Plan implementation, index of, 116
Plans, Development, 7, 224-5
Plowden Report, 195
Policy:
  general, 44, 58
  objectives, 7-8
  use of indicators in statement, 7
Political cleavage, Index of, 100
Political development index, 85, 88
Pragmatic reduction, 34
Preferences:
  for education, 28, 148
  for other facilities, 29
Pre-school, 151
Processes:
  definition of, 25-7
  indicators of, 149-65, 192-8, 234-40, 247 ff, 269-73
Profiles, 7, 218-19, 260-2

Promotion rates, *see* Transition rates
Provincial education systems, *see* Regional education systems
Puerto Rico, 13
Pupil-teacher ratio, 190

Qualitative features of systems, 44, 54-5, 192
Questionnaires, 63, 85-7, 148, 172-4

Race, 40, 91, 95-6, 134
Rates of return, 178-80
Ratio measurement, 20, 151
Ratio scales, 47-50, 78-9, 118, 122, 132
Regional education systems, 30, 40, 42, 61, 89, 160-2, 189-204, 209-11, 234-40, 269-72
Regression analysis, 78, 119-20, 131-3, 185-6, 201
Reliability, 31, 38, 49, 59-69, 190, 195, 201
Religion, 12, 47
Repetition rates, *see* Transition rates
Replication analyses, 37, 151, 192, 285
Reports, 100, 205-7
Representative indices, 15-16, 34, 70
Riot severity index, 134
Rwanda, 79-81, 208

Samples, 62-3, 173-4, 199-200
Satisfaction with education, 29, 86-7, 107-8, 123, 172-4
Saudi Arabia, 97-8, 110-14, 155, 208, 224-34
Schools, relative availability of (index), 152, 235-7, 241
Secondary sources of data, *see* Indirect measurement
Self-referencing, 17-19, 20, 23, 101-15, 125, 207, 242-3
Social Indicators, 34, 220
Social mobility, 259
Societal characteristics, 246-7
South America, 18, 79-81, 134-5, 162-3, 251
Spain, 268
Special education, 48, 57, 151, 227
Specialization of graduates – index, 171
Squares, *see* Cubes
Stability of definition, 124
    see also Replication analyses
Standard deviation, 22, 79-81, 83-5
Standard scores, 82-5, 128, 152, 178
Statistician, definition of, 81
Stocks, 21
Structural differentiation, index of, 42, 134-5, 162-3
Student places – relative availability index, 152-3, 226-8, 235-7
Subtraction, *see* Addition
Surinam, 19
Surveys, 33, 62-6, 148, 172-4, 198-9, 244
Synchronic, *see* Norm-referencing

T scores, *see* Standard scores
Tables (data), 206
Tanzania, 13
Targeting, 8, 12, 268
Teachers:
    qualification of, 83, 96, 100-1, 158-9, 238-40, 259
    supply of, 176-7
Technical education, 154, 220, 230-1, 237, 250
Test scores, 43, 66, 91, 198
Thailand, 90, 119, 250
Theoretical definition of indicators, 34-6
Theory, 5-6, 24, 33, 37, 41-2, 256, 257
Transition rates, 109-14
Turkey, 253
Typologies, 12-14, 257 ff, 297-301

UNDP, 13
Unemployed, education level of, 176
Unesco, 13, 21, 51
Unicef, 51
United Kingdom, *see* England
United Nations, 21
Urbanization, 31, 38, 100
Uruguay, 19, 117
USA, 13, 40, 67, 96, 107-8, 170, 172-3, 189, 196-9, 202-3, 220, 249
USSR, 56, 220

Validity, 32, 38, 51, 55-8, 69-70, 131, 190
Venezuela, 19, 216, 249
Vocational education, *see* Technical education

Wales, 157
Weights, 35-6, 76-7, 80-1, 99, 104-8, 169
Wisconsin, 196-8, 202-3, 209
World Bank, 13, 51

Years of schooling, expected number, 11, 43, 91, 109-12, 163, 166, 170, 232-4

Yugoslavia, 119, 267

z scores, *see* Standard scores
Zambia, 79-81, 208